Optical Wavefront Prediction with Reservoir Computing

Stephen J. Weddell, M.App.Sc. (Eng.)

Department of Electrical and Computer Engineering

A thesis presented for the degree of
Doctor of Philosophy

University of Canterbury
Christchurch, New Zealand
January 2010

Dedicated to my family: Sarah, Thomas, and Margie.

A man said to the Universe: 'Sir, I exist.' 'However', replied the Universe, 'The fact has not instilled in me a sense of obligation'.

Stephen Crane

Abstract

Over the last four decades there has been considerable research in the improvement of imaging exo-atmospheric objects through air turbulence from ground-based instruments. Whilst such research was initially motivated for military purposes, the benefits to the astronomical community have been significant. A key topic in this research is isoplanatism. The isoplanatic angle is an angular limit that separates two point-source objects, where if independent measurements of wavefront perturbations were obtained from each source, the wavefront distortion would be considered equivalent. In classical adaptive optics, perturbations from a point-source reference, such as a bright, natural guide star, are used to partially negate perturbations distorting an image of a fainter, nearby science object.

Various techniques, such as atmospheric tomography, maximum *a posteriori* (MAP), and parameterised modelling, have been used to estimate wavefront perturbations when the distortion function is spatially variant, i.e., angular separations exceed the isoplanatic angle, θ_0, where $\theta_0 \approx 10\,\mu$rad for mild distortion at visual wavelengths. However, the effectiveness of such techniques is also dependent on knowledge *a priori* of turbulence profiles and configuration data.

This dissertation describes a new method used to estimate the eigenvalues that comprise wavefront perturbations over a wide, spatial field. To help reduce dependency on prior knowledge for specific configurations, machine learning is used with a recurrent neural network trained using *a posteriori* wavefront ensembles from multiple point-source objects. Using a spatiotemporal framework for prediction, the eigenvalues, in terms of Zernike polynomials, are used to reconstruct the spatially-variant, point spread function (SVPSF) for image restoration. The overall requirement is to counter the adverse effects of atmospheric turbulence on the images of extended astronomical objects. The method outlined in this thesis combines optical wavefront sensing using multiple natural guide stars, with a reservoir-based, artificial neural network. The network is used to predict aberrations caused by atmospheric turbulence that degrade the images of faint science objects.

A modified geometric wavefront sensor was used to simultaneously measure phase perturbations from multiple, point-source reference objects in the pupil. A specialised recurrent neural network (RNN) was used to learn the spatiotemporal effects of phase perturbations measured from several source references. Modal expansions, in terms of Zernike coefficients, were used to build time-series ensembles that defined wavefront maps of point-source reference objects. The ensembles were used to firstly train an RNN by applying a spatiotemporal training algorithm, and secondly, new data ensembles presented to the trained RNN were used to estimate the wavefront map of science objects over a wide field. Both simulations and experiments were used to evaluate this method.

The results of this study showed that by employing three or more source references over an angular separation of 24 μrad from a target, and given mild turbulence with Fried coherence length of 20 cm, the normalised mean squared error of low-order Zernike modes could be estimated to within 0.086.

A key benefit in estimating phase perturbations using a time-series of short exposure point-spread functions (PSFs) is that it is then possible to determine the long exposure PSF. Based on the summation of successive, corrected, short-exposure frames, high resolution images of the science object can be obtained. The method was shown to predict a contiguous series of short exposure aberrations, as a phase screen was moved over a simulated aperture. By qualifying temporal decorrelation of atmospheric turbulence, in terms of Taylor's hypothesis, long exposure estimates of the PSF were obtained.

Acknowledgements

Upon reflection of those who have helped me with this research, I am extremely grateful to my supervisor, Dr. Russ Webb, who continues to challenge me on different perspectives related to my work. I have learnt a great deal from Russ and have appreciated his style of supervision. Also, I would like to thank my assistant supervisor, Prof. Phil Bones, for his consideration with teaching loads and enthusiastic support throughout this work.

I am indebted to my family for allowing me the opportunity to pursue this obsession. After many late nights working on simulations, papers, and this dissertation, an occasional sanity check is extremely helpful; my family, Sarah, Thomas, and Margie have provided this, and much more, with their love, patience, and understanding.

During the course of this work, several colleagues, postgraduates, and students have been instrumental in helping pursue an idea or concept, either at the whiteboard, or over a cup of coffee. These include Prof. Des Taylor for enlightening me on the inner workings of journal associations and the submission process, Assoc. Prof. Peter Smith for his course on Applied Random Processes and related discussions, Mark Noonchester, Tom Cronje, and Ciaran Moore, for their offers of bike rides to clear the head and coffee to endure those long nights, and my project students, particularly Ricky Tiong, for his design, construction, and subsequent publication of a hot-air turbulence generator for the Optics Laboratory - thank you one and all.

Early in my research I appreciated the many discussions in tutorials or laboratories with Dr. Yong Chew. As an active researcher in the field, Yong was able to clarify many challenging concepts regarding optics and imaging, and their application to adaptive optics. During the course of my experimental studies, Dr. Judy Mohr has been a great asset. Our combined observation runs at Mt. John Observatory provided me with the confidence to continue to acquire empirical data for use in my research. Lastly, I would like to thank the team at Mt. John, specifically Pat Kilmartin and Alan Gilmore, who have made me feel very

welcome during my short stays over the years, and for allowing me to tap their vast pool of astronomical knowledge.

Contents

Abstract **v**

Acknowledgements **vii**

Contents **ix**

Preface **xvii**

0.1 Motivation xviii

0.2 Original contributions xviii

0.3 Supporting publications xx

Glossary **xxi**

0.4 Notation xxi

0.5 Abbreviations xxiii

1 Introduction **1**

1.1 Natural and artificial guide stars 4

1.2 Atmospheric turbulence profiling 5

1.3 Modelling vs. machine learning 6

1.4 Thesis organisation 8

2 Mathematical Preliminaries **11**

2.1 Notation 12

2.1.1 Vectors and matrices 12

2.1.2 Coordinate spaces 12

2.2 Random variables and random processes 13

2.2.1 Random variables 13

2.2.2 Random signals and random processes 16

2.2.3 Moments 17

2.2.4 Performance metrics 18

2.2.5 Noise 20

2.3 Linear methods 21

2.3.1 Linear regression 22

2.3.2 Singular value decomposition 23

2.3.3 Pseudoinverse 23

2.3.4 Parameter estimation 24

2.4 Regularisation 25

3 Background **27**

3.1 Geometrical optics 28

3.1.1 Optical path lengths and differences 28

3.1.2 The optical wavefront 29

3.2 Fourier optics 30

3.2.1 Diffraction 30
3.2.2 Diffraction in the generalised pupil 32
3.3 Modal expansions 33
3.3.1 Zernike polynomials 34
3.4 Intensity and spectral transfer functions 36
3.4.1 The point spread function 36
3.4.2 The optical and modulation transfer functions 39
3.5 Coherent and incoherent imaging 40
3.6 Long and short exposure imaging 42
3.6.1 Long exposure transfer functions 43
3.6.2 Short exposure transfer functions 44
3.7 The Image Model 45
3.7.1 The spatiotemporal PSF 48
3.7.2 Spatial variance of the anisoplanatic PSF 49

4 Adaptive Optics 51

4.1 Atmospheric turbulence 55
4.1.1 Characterising air turbulence 55
4.1.2 The Kolmogorov turbulence model 56
4.1.3 Temporal decorrelation 58
4.2 Time series analysis of Zernike modes 59
4.2.1 Background 60
4.2.2 Acquisition 61

4.2.3 Analysis 61
4.3 Imaging through turbulence 65
4.3.1 Anisoplanatism 65
4.3.2 The Taylor frozen turbulence hypothesis 67
4.3.3 Residual phase error 68
4.4 Wavefront sensors 74
4.4.1 Background 74
4.4.2 The Curvature WFS 76
4.4.3 The Geometric WFS 79
4.4.4 Simultaneous wavefront measurements from multiple source beacons 80
4.4.5 Limitations of extended curvature and geometric sensors 82
4.5 Deconvolution from wavefront sensing 83
4.6 Estimating the anisoplanatic PSF 85
4.6.1 Wavefront estimation using multiple source beacons 86
4.6.2 The AO-compensated anisoplanatic PSF 91
4.7 Qualifying wavefront deformation in the metapupil 92
4.7.1 Near ground turbulence effects on anisoplanatic imaging 93
4.7.2 Wavefront sensing constraints on near ground turbulence 95
4.7.3 Review of Taylor's frozen hypothesis 97
4.7.4 Quantifying Taylor's hypothesis for wide-field imaging 99

5 Reservoir Computing **103**

5.1 Background 104

5.2 Design considerations . . . 104
5.3 Echo state networks . . . 107
5.3.1 Introduction . . . 107
5.3.2 Architecture . . . 107
5.3.3 Training . . . 109
5.3.4 Normalisation and activation functions . . . 110
5.4 Echo state network optimisations . . . 112
5.4.1 Design parameters . . . 113
5.4.2 Performance evaluation . . . 117
5.5 Noise and memory considerations . . . 119
5.6 Prediction vs. estimation . . . 120

6 Simulations **121**

6.1 System model . . . 122
6.2 The propagation model . . . 124
6.2.1 Overview . . . 125
6.2.2 Propagation model parameters . . . 127
6.2.3 Phase screen generation . . . 128
6.2.4 Plane wave propagation through turbulence . . . 129
6.2.5 Phase estimation and imaging . . . 132
6.2.6 Sampling requirements . . . 135
6.2.7 Operational summary . . . 138
6.2.8 Data structures . . . 139

6.3 The prediction model 142
6.3.1 Prediction model parameters 143
6.3.2 Spatiotemporal training 143
6.3.3 Data generation algorithms 147
6.3.4 Spatiotemporal prediction algorithm 150
6.3.5 Verification model 154

7 The ESN Method 157

7.1 General test procedure and parameter sets 158
7.2 Prediction by field angle and sources 159
7.3 Small field angles 162
7.4 Prediction by Fried's parameter and field angle 162
7.5 Performance evaluation with CCD noise constraints 165
7.5.1 Test procedure 167
7.5.2 Results 167
7.6 Performance evaluation with photon noise constraints 169
7.6.1 Test procedure 171
7.6.2 Results 172
7.7 Network performance with combined noise effects 174
7.7.1 Input data dependency 177
7.8 Performance summary 180

8 Conclusions and Future Research 183

8.1 Conclusions 183

8.2 Future research 185

A Restoration of Astronomical Images 189

A.1 Introduction 190

A.2 Degrees of freedom 190

A.2.1 PSF spatial constraints 191

A.3 PSF error maps 193

A.4 Open loop imaging model 195

A.5 Image restoration 196

A.5.1 Restoration algorithms 196

A.5.2 Isoplanatic blocks 197

A.5.3 The anisotropic image 198

A.5.4 Block processing 199

A.6 Restoration example 202

A.6.1 Procedure 202

A.6.2 Restoration 202

A.6.3 Results 204

A.7 Summary 205

B Laboratory Equipment and Experimentation 207

B.1 Equipment 207

B.2 Laboratory 210

B.2.1 Test bench 210

B.2.2 Laboratory procedure 210

C Observational Data **215**

D Derivation of the Continuous and Discrete Point Spread Function **221**

E Summary of Mt. John University Observatory Wind Velocity Models **225**

References **227**

Preface

After completing the Masters degree at Curtin University of Technology in Perth, I worked for several years before returned to academia in 2000. Soon after accepting an application engineering position in the Department of Electrical and Electrical Engineering at the University of Canterbury, I was offered additional responsibilities, including a role as course coordinator in computer hardware engineering. I continue to lecture this course and have developed several new courses along the way. However, as much as I have enjoyed lecturing, the urge to continue postgraduate study was always present. I, like the majority of my academic colleagues, strongly believe that teaching and research are inseparable.

Background work on this dissertation began many years prior to the commencement of formal study. From an early age, I held a strong interest and fascination in astronomy. Combining this interest with engineering principles and knowledge of science has, for me, fused together several great passions resulting in a rewarding journey.

More specifically, the anisoplanatic problem has always fascinated me, especially from a signal processing perspective. The current techniques available to reduce the effects of anisoplanatism are effective, to a degree, however the implementation of a system that can learn spatiotemporal effects for estimation of phase perturbations over a wide field was a compelling goal. In addition to the theory, practical aspects to this research were considered essential. This included field studies at Mt. John University Observatory near Lake Tekapo in New Zealand, typically over term breaks between 2005 and 2009. An appreciation of the practical constraints in acquiring field data, as distinct from defining simulation parameters, was quickly learnt. These lessons were incorporated in the model discussed in Chapter 6 of this dissertation.

In order to formulate the results contained in this work, specialised knowledge was required; such knowledge spanning over several major disciplines. Previous research in related topics such as wavefront sensing, neural networks, and inverse problems in astronomical imaging,

principally from fellow UC researchers, has been a valuable resource, both in terms of background knowledge and inspiration.

0.1 Motivation

The development of an architecture that supports a model-less system is appealing. A model-less system is not limited to a predetermined transfer function; new sequences can be learned through training to produce desired outputs. The adaptation of new recurrent networks that simplify training and hardware implementation, such as reservoir computing, is a strong motivation and offers significant benefits in application to the field to adaptive optics and astronomical imaging. Lastly, some extensions to this work may provide a foundation for application in other related fields of science and engineering.

0.2 Original contributions

The work completed for this research, and the results presented in this thesis, comprise a significant proportion of original contributions. The Geometric wavefront sensor (WFS), originally used for single-source wavefront sensing [19], provided a basis for new research in sensing wavefronts from multiple source beacons and targets over a wide field. Deconvolution from wavefront sensing (DWFS) was used, both in simulation and laboratory work, employing a Geometric sensor over a wide field-of-view (FOV). A three camera configuration, two slightly defocused high frame-rate cameras used for Geometric sensing, and one focused target camera for image reconstruction, provided a versatile simulation and reliable observational platform. Perturbations from multiple sources and targets were evaluated simultaneously and formed the basis of time-series data that was used for training a recurrent neural network (RNN) for prediction of the anisoplanatic point spread function (PSF), and subsequent restoration of target point source objects.

Taylor's frozen turbulence hypothesis is an important consideration for application of the ESN method. In testing the spatial extent of Taylor's hypothesis using empirical data, a method was developed to quantify Taylor's hypothesis, in terms of atmospheric decorrelation over wide, spatial fields.

The integration of reservoir computing for prediction of the spatially variant point spread function (SVPSF) forms a significant portion of new work. Simplified training requirements, supported by an echo state network (ESN), provided the basis for predicting the

anisoplanatic PSF using a wavefront propagation model. A spatiotemporal image model was developed using angular separations between multiple source beacons and a single target. The relationship between these original research contributions is outlined in Figure 1.

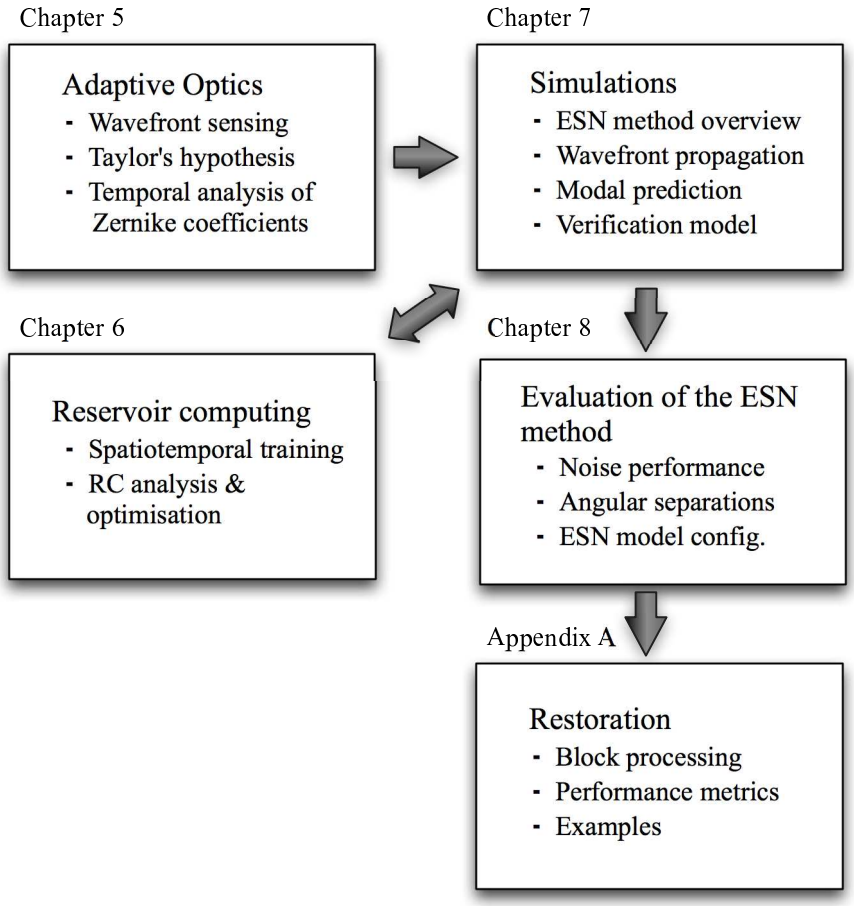

Figure 1 Original contributions.

Extensive wavefront simulations for modal predictions were conducted to refine ESN architecture. This is represented in Figure 1 by a bidirectional arrow for Chapters 5 and 6. By analysing parameters used for the spatiotemporal image model, in conjunction with the ESN method proposed for this study, a new technique has been developed for the acquisition of aberrated wavefront data for network training and testing.

Complementary to this research topic, subsequent contributions have been made. For example, in assessing the quality of anisoplanatic images restored using the ESN method,

distortions along the boundaries of adjoining sub-images, or more specifically, isoplanatic blocks, were noted. The restoration of anisoplanatic images using an enhanced block processing method, is detailed in Appendix A.

0.3 Supporting publications

Journal and conference papers prepared and published as a result of research detailed in this thesis, are listed below in order of publication.

"Reservoir Computing for Prediction of the Spatially-Variant Point Spread Function", Weddell, S. J. and Webb, R. Y., Selected Topics in Signal Processing, IEEE Journal of, vol.2, no.5, pp. 624-634, Oct. 2008

"The restoration of extended astronomical images using the spatially-variant point spread function", Weddell, S. J. and Webb, R. Y., in Proceedings of Image and Vision Computing New Zealand, 2008 (IVCNZ'08), K. Irie (Editor), November 2008, IEEE Press, pp 1-6.

"A neural network architecture for the reconstruction of turbulence degraded point spread functions", Weddell, S. J. and Webb, R. Y., in Proceedings of Image and Vision Computing NZ 2007 (IVCNZ'07), M.J. Cree (Editor), Hamilton, December 2007, pp 103-108.

"Dynamic artificial neural networks for centroid prediction in astronomy", Weddell, S. J. and Webb, R. Y., Sixth International Conference, Hybrid Intelligent Systems (HIS'06) and 4th Conference on Neuro-Computing and Evolving Intelligence (NCEI'06), I. Havukkala (Editor), Auckland, December 2006, IEEE Press, pp 68-72.

"Data Preprocessing on Sequential Data for Improved Astronomical Imaging", Weddell, S. J. and Webb, R. Y., in Proceedings of Image and Vision Computing New Zealand 2005 (IVCNZ'05), B. McCane (Editor), Dunedin, December 2005.

Glossary

0.4 Notation

j	$\sqrt{-1}$
a_i	The i^{th} modal expansion coefficient
$\mathbf{A}^{-1}$	Inverse of matrix $\mathbf{A}$
$\mathbf{A}^T$	Transpose of matrix A
C_n^2	Refractive index structure constant
$\delta(\mathbf{x})$	Dirac delta function
d	Distance
D	Telescope aperture dimension
$D(r)$	Structure function
ε^2	Mean squared error (MSE)
ε_R^2	Mean squared residual wavefront error
$f(\cdot)$	An object
$\mathrm{FT}\{\cdot\}$	Fourier transform operation
$g(\cdot)$	Measured distorted object
h	Height above the optical aperture (m)
$h(\cdot)$	Point spread function
$H(\cdot)$	Optical transfer function
k	Optical wave number, $2\pi/\lambda$
$M(\cdot)$	Modulation transfer function

$\eta(\cdot)$	Measured noise
θ_0	Isoplanatic angle
r_0	Fried parameter
$\mathbf{r}$	Vector distance in 3D space
λ	Optical wavelength (m)
λ_i	The i^{th} eigenvalues of a matrix
λ_j	The j^{th} Lyapunov exponent
m_v	Apparent magnitude
$\lvert\cdot\rvert$	Magnitude of
$\lVert\cdot\rVert$	Norm of
$\lVert\cdot\rVert_n$	n^{th} power norm of
$\rho(\cdot)$	Spectral radius
Σ	Summation operator
$P(\cdot)$	Pupil function
$\mathscr{P}(\cdot)$	Generalised pupil function
Π	Product operator
$\odot$	Convolution operator
$*$	Correlation operator
μ	Mean of a random variable
$\phi(\cdot)$	Phase in the pupil
σ	Standard deviation
σ^2	Variance
$\psi(\cdot)$	Wavefront phase function
S_i	The i^{th} reference source beacon (see NGS and LGS)
T_j	The j^{th} target or science object
$U(\cdot)$	Complex amplitude of an optical wave
$V_{\text{wind}}(h)$	Wind velocity as a function of height, h

$\mathbf{W}$	ANN weight matrix
$W(\cdot)$	Wavefront intensity aberration function
W_j	Metapupil for the j^{th} layer
$\mathbf{x}$	2D vector
z	Distance along optical path
Z_i	The i^{th} Zernike polynomial (as adopted by Noll [98])
$\{\cdot\}$	Set of elements
$\langle\cdot\rangle$	ensemble average

0.5 Abbreviations

AMSE	Averaged mean squared error
ANN	Artificial neural network
AO	Adaptive optics
CCD	Charge coupled device
CAO	Classical adaptive optics
DR	Dynamic reservoir
DWFS	Deconvolution from wavefront sensing
ESN	Echo state network
fBm	Fractional Brownian motion
FFN	Feedforward network
FFT	Fast Fourier transform
FOV	Field of view
IID	Independent identically distributed data
IO	Input and output
LE	Long exposure
LGS	Laser guide star
LP	Linear predictor

MAP	Maximum *a posteriori*
MMSE	Minimum mean squared error
MJUO	Mount John University Observatory
MSE	Mean squared error
NMSE	Normalised mean squared error
NGS	Natural guide star
OPL	Optical path length
OTF	Optical transfer function
pdf	probability density function
PE	Processing element
PSF	Point spread function
ROI	Region of interest
RNN	Recurrent artificial neural network
SE	Short exposure
SNR	Signal to noise ratio
SIPSF	Spatially invariant point spread function
SR	Strehl ratio
SVD	Singular value decomposition
SVPSF	Spatially variant point spread function
STM	Spatial training matrix
WFS	Wavefront sensor
ZCV	Zernike coefficient vectors

Chapter 1

Introduction

Photons, emitted from stars, travel uninhibited over billions of kilometres through space only to be viewed on Earth many years later as fluctuations of light, their former brilliance masked by a relatively thin layer of atmosphere. Instruments, such as optical telescopes, provide the capability to look back in time and approach the threshold of the birth of our Universe. However until very recently, our view and understanding of the universe was severely limited, irrespective of the size or type of ground-based optical instrumentation. Atmospheric turbulence reduces the resolution of optical telescopes; quantitatively, this can be in excess of an order of magnitude. Recent technology, introduced by research in adaptive optics [134], has all but removed these limitations under ideal conditions. The work outlined in this dissertation is offered as a further contribution to the vast wealth of existing knowledge that comprises the field of adaptive optics and the discipline of signal processing.

Long exposure imaging through atmospheric turbulence results in the distortion of faint astronomical science objects due to the aberration of effectively planar wave fronts. Wavefront sensors provide indirect measurement of wave fronts from a reference beacon, however two constraints are imposed by this method. Firstly, a reference beacon needs to provide sufficient irradiance to maximise the signal to noise ratio (SNR) from the wavefront sensor. Secondly, wavefront phase measurements from a reference beacon should, ideally, be identical to wavefront phase (hypothetically) measured from the science object. The limit imposed on effectively measuring the same phase from two objects separated by an angle of θ, is known as the isoplanatic angle, θ_0, where $\theta \leq \theta_0$.

Given the ideal case of the existence of a bright reference beacon within the isoplanatic

patch of a target object, the spatially-invariant (on-axis) point spread function (PSF) can be recovered. The PSF is referred to as the 2D impulse response function and since it comprises the distortion effects of the atmosphere, widely used deconvolution algorithms can employ the PSF to partially restore an image of the original target object. However, in the majority of situations concerning astronomical imaging, such an ideal case is rarely presented. There is an insufficient number of bright, naturally occurring reference beacons, often referred to as natural guide stars (NGSs), within isoplanatic patches of target objects to be effectively used as reference sources [90]. The development of artificially created source beacons, referred to in the literature as a laser guide star (LGS), have partially addressed this problem, and an overview of references sources is given in Section 1.1.

In order to compensate for the majority of situations that are technically more challenging than the ideal example discussed above, various methods, such as maximum *a posteriori*-based estimation [32], and modal tomography [109], have been developed over the last decade. These methods employ multiple reference beacons over a wide field-of-view (FOV) for wavefront recovery within anisoplanatic regions. In this thesis, a similar approach is taken, however a specialised recurrent neural network (RNN) is used to estimate phase over a wide FOV and over a succession of short exposure images.

A somewhat imaginative yet intuitive example of the atmospheric imaging problem can be described as follows [134]. Consider a multitude of different sized lenses, contained within a certain boundary or layer, surrounding the Earth. Each lens has the ability to move over a telescope aperture, and within the FOV of the optical instrument. The size of each lens can be defined in terms of a coherent cell length, r_0 [30], where values for each cell can range between 2 cm to 300 cm. Group of cells can move at different velocities and directions within a bounded layer, but are generally propelled in a uniform direction by prevailing winds. A model that describes multiple layers of such bounded cells is often referred to as the *thin layer* turbulence model. Such models are typically based on two or three layers of thin, turbulent slices of atmosphere, contained within specific boundary regions; these include the surface and boundary layers (0-1 km and 1-10 km above ground, respectively), in addition to the troposphere (10-25 km) [49]. Such a model, incorporating a telescope to image a single point source object at infinity, is shown in Figure 1.1.

The restoration of astronomical images, distorted by the effects of imaging through the turbulent atmosphere, is based on mathematical principles, applied through digital image and digital signal processing. The real-time restoration of such images can be achieved using adaptive optics (AO), where the optical path of an imaging system is altered by applying

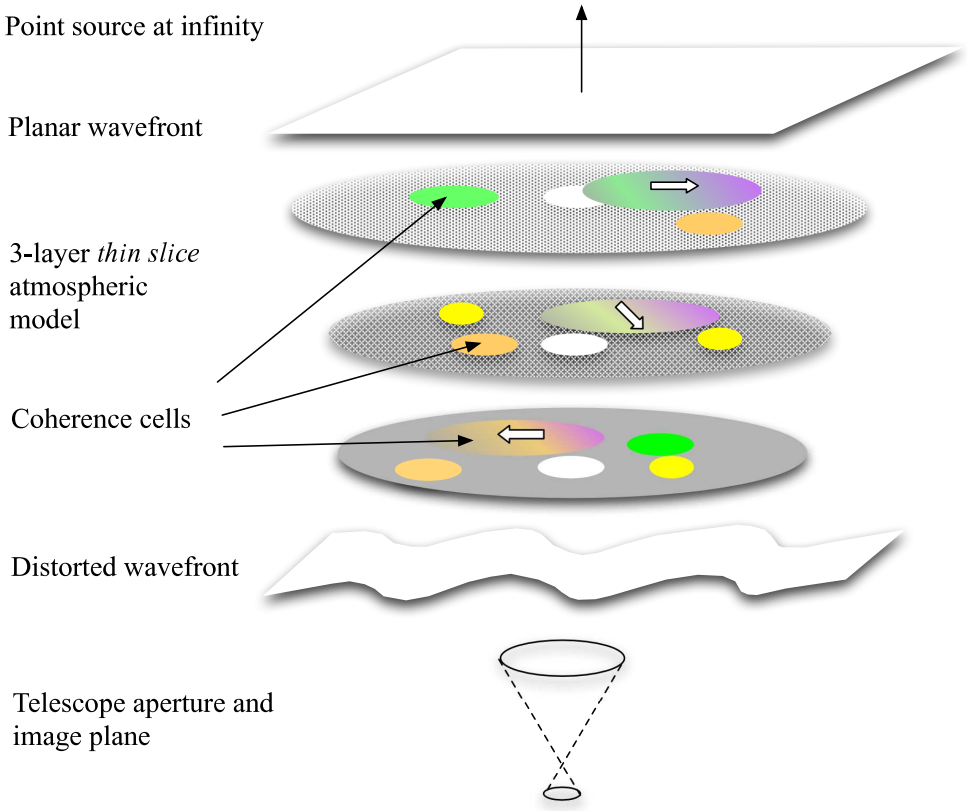

Figure 1.1 A three layer *thin-slice* turbulence model.

the conjugate of the measured wavefront to a deformable mirror [115], thus closing the control loop. As AO systems are extended, greater demands are placed on the control system. Open-loop systems, such as deconvolution from wavefront sensing (DFWS) [117], provide the basis for this study by allowing an estimate of a target object to be made using aberration data from multiple reference beacons. However, to meet the requirements for wide field astronomical imaging, i.e., to minimise the error between an estimated wavefront perturbation and the actual wavefront perturbation over a wide FOV, an efficient method for the prediction of the spatially-variant PSF is required.

Equipment used in this research, such as CCD cameras, optics, and computing, is based on off-the-shelf componentry. Justification of this approach is primarily that of convenience. Researchers working in similar respective fields can duplicate the methods developed in this study, and presumably reproduce the same results with minimal outlay. An example of this approach is the use of high-speed CCD cameras as an alternative to photon counting avalanche diodes typically used for curvature wavefront sensing [10]. Data acquired from a

1-m class optical telescope at Mt. John University Observatory (MJUO) near Lake Tekapo New Zealand, and from an Optical Laboratory emulating the effects of turbulence on multiple planar wavefronts, have been used for a time series analysis detailed in this thesis. The results from this analysis facilitated optimisation of the method outlined in this dissertation.

This thesis discusses several major areas in adaptive optics where machine learning has provided improvements, in terms of the prediction and classification of wavefronts. Specifically, Zernike coefficients to the 20^{th} order are predicted to within 5% of mean squared error (MSE) for anisoplanatic angles of 24 μradians using a trained, recurrent neural network. A new method to image multiple source objects over a wide FOV for deconvolution from wavefront sensing (DFWFs) is presented. These results are then combined with a new technique to interpolate the spatially variant point spread function (SVPSF), based on modal decomposition for reconstruction of extended objects using existing deconvolution algorithms.

1.1 Natural and artificial guide stars

In the preceding section, the concept of using a bright neighbouring star as a reference to estimate wavefront aberrations from a faint target object was introduced. This concept is an important prerequisite for the method detailed in this thesis. As a result, a brief background on guide stars is given in this section.

Three problems exist concerning the employment of guide stars to estimate target wavefronts:

1. The sparsity of bright natural stars: It has been reported that the probability of a 10^{th} magnitude star or lower (brighter), being within 20-40 arcseconds of an astronomical science object, is less than 0.15% [90]. The sparsity of bright natural stars promotes the use of fainter stars as references.

2. Detector noise: Detectors used to measure optical wavefronts collect less photons from faint stars than from bright stars. To compensate photon starvation, high gain sensors are used. However CCD image sensors suffer from read noise and this results in low SNR. The use of avalanche photodiode (APD) image arrays significantly reduces read nose, however photon noise remains a major obstacle in obtaining accurate wavefront measurements.

3. Angular anisoplanatism: As the angular separation between a guide star and target object increases, so does the wavefront error. This error can be measured in terms of wavefront variance, σ^2, for any angle θ, and can be expressed as, $\langle\sigma_\theta^2\rangle = (\theta/\theta_0)^{5/3}$, where θ_0 is the isoplanatic angle. This condition is known as *angular* anisoplanatism [30], and $\langle\sigma_\theta^2\rangle$ is the anisoplanatic error [48].

In order to overcome Item-1, an artificial, or laser guide star (LGS) can be created by using a finely tuned laser beam to excite naturally occurring atomic sodium at an altitude of approximately 90 km [25]. The intensity of a LGS can be regulated sufficiently to address Item-2. Also, the creation of a LGS means that they can be placed within a region 30 arcseconds of the science object [103], thereby addressing Item-3.

However, the application of LGSs in AO is not without limitations. As outlined by McGuire *et al.* [90], the projection of the artificial guide star from the ground means that the upwardly propagating laser beam samples the same turbulence as the returned, i.e., backscattered laser light. Each subaperture used for wavefront sensing will only be able to measure wavefront tilt relative to the upwardly projected beam, which is itself experiencing a global deflection relative to the science object. This limiting condition is referred to as *focal* anisoplanatism, or the *cone effect* [134].

Countering the effects of focal anisoplanatism is not addressed in this thesis. A comprehensive analysis of anisoplanatic effects is given by Fried [30]; additionally, and related to this research, Tallon and Foy propose a method to minimise focal anisoplanatism [126]. This thesis proposed a new method to reduce the effects of angular anisoplanatism.

In summary, classical adaptive optics uses wavefront aberrations measured from bright source references, such as NGSs or LGSs, separated by small field angles from a target. In classical AO systems, the conjugate of these wavefront measurements is applied to the optical path of the system to correct atmospheric distortion and improve the resolution of optical telescopes. The sparsity of bright sources has promoted the use of LGSs, however their application is limited by focal anisoplanatism.

1.2 Atmospheric turbulence profiling

A relatively simple method has been developed to measure and profile turbulence in the atmosphere. This technique, introduced by Vernin and Roddier in 1973 [138], is known as

scintillation detection and ranging (SCIDAR). The image of two stars that form a binary pair are observed. The overlapping pupil images are recorded for later analysis to determine the vertical profile of the atmospheric turbulence. The processing of this information is used to profile AO sites.

An alternative turbulence profiling method is slope detection and ranging (SLODAR). The local wavefront tilt is typically measured using a Hartmann-Shack[1](H-S) wavefront sensor to provide slope measurement [145]. Light from a binary star is employed, of angular separation, θ, to determine the correlation of instantaneous wavefront slopes between the stars.

Control systems form an essential part of an Adaptive Optics installation. The results from SCIDAR runs can be used to precisely calibrate control systems to specific altitudes of atmospheric turbulence and form part of a site profile. Typically, there are between two to three turbulent, atmospheric layers that surround the Earth. The introduction of Chapter 4 discusses how the effects of atmospheric turbulence can be corrected by Adaptive Optics. However, *a priori* parameters concerning atmospheric conditions at specific sites provide the bounds within which an AO system can operate.

In this thesis, an atmospheric model of MJUO is used in both simulations and observations [91]. The ability to remotely record atmospheric turbulence has been simplified considerably with the construction of a portable optical breadboard for SCIDAR application [63]. This system has been successfully used to characterise atmospheric turbulence at MJUO [91], and has been modified for the purposes of this study.

1.3 Modelling vs. machine learning

An important distinction in this research is made between a system that uses a model to determine a solution to a problem, and a system that effectively learns a method of mapping input signals to outputs to find a solution. In terms of adaptive, signal processing applications including estimation, the former requires *a priori* knowledge of a specific definition of the problem, in the form of differential or difference equations. State-space methods are a natural extension to formal modelling techniques, and are used to describe a dynamical system. The latter encompasses nodal-based architectures that, when provided with a system

[1]The majority of references to this wavefront sensor reverse the order of the inventors to that shown in this thesis, i.e., the Shack-Hartmann sensor, however Hartmann is placed first in recognition of his invention in 1900, and secondly, Shack's enhancements to the Hartmann WFS in the 1960's.

that can learn to emulate any multivariate linear and non-linear function [50], effectively formulates its own model through training. The term used for such model-less systems is *machine learning* and is of significant importance in this thesis.

Machine learning can be divided into two major classifications, generative and discriminative learning [56]. An interim group, conditional learning, is classified as a mixture of both generative and discriminative learning. Both primary learning classifications have been used to estimate phase perturbations resulting from wavefront aberrations due to the effects of atmospheric turbulence [32], [22]. Each of these methods are discussed in Section 4.6. However, Table 1.1 provides an overview of alternative methods of estimation.

Table 1.1 Classifications in machine learning; adapted from Jebara [56].

	Classifications		
	Generative	**Conditional**	**Discriminative**
Local	Maximum likelihood (ML)	Maximum conditional likelihood	Empirical risk
Local and prior	Maximum *a posteriori* (MAP)	Maximum conditional *a posteriori*	Artificial neural networks (ANNs).
Model averaging	Bayesian inference	Conditional Bayesian inference	Maximum entropy (ME)

The generative and discriminative classifications listed in Table 1.1 are discussed in this thesis. The generative approach has been used to estimate the effects of turbulence by maximising *a posteriori* the outcome of the process. This method is discussed in Section 4.6. Other generative and discriminative methods, such as maximum likelihood and maximum entropy used for astronomical image restoration, are discussed by Leung [75].

The introduction to Chapter 5 provides details on recurrent neural networks used for this study, i.e., a discriminative learning approach. Conditional forms of Bayesian inference for estimation of a parameter value, maximised with and without a prior, are detailed in Jebara [56].

1.4 Thesis organisation

This dissertation can be broadly divided into five major topics, each spanning one or more chapters. These include:

1. Mathematical preliminaries and background on optical principles and image models, used throughout this thesis.

2. Adaptive optics, including details on wavefront sensor extensions developed for simulations and field observations; a temporal analysis of low-order Zernike coefficients; discussion on the spatial constraints of Taylor's hypothesis; a summary of current methods used to estimate the anisoplanatic point spread function.

3. An overview of reservoir computing; architectural details on the adaptation of a machine learning system for prediction of modal wavefront aberrations; spatiotemporal training algorithm; system optimisation.

4. Development of a propagation model for simulation; results from simulations, both in terms of wavefront propagation and estimation of modal wavefront perturbations; performance analysis with additive and photon noise.

5. Practical application of the results of this study.

Notational and mathematical foundations used in the development of this thesis are presented in Chapter 2. Background material on optics and diffraction approximations are summarised in Chapter 3, including a section on Zernike polynomial modal expansions that form the basis of data ensembles used for this study.

Recovery of the space variant point spread function is used to define image degradation over a wide field-of-view. Chapter 3 also reviews specific image models that define both short exposure and long exposure transfer functions for astronomical imaging. Temporal information is incorporated in such functions resulting in the formulation of a spatiotemporal image model.

Adaptive optics is a broad technology, founded on signal processing and scientific principles. However, Chapter 4 presents a focused subset of these principals to support the contributions detailed in this dissertation. For example, the Taylor *frozen* turbulence hypothesis is outlined and qualified, both temporally and spatially, over an extended field

using observational data. Existing research on curvature sensing is expanded, facilitating both modal and zonal operation for multiple guide star wavefront sensing. An overview of existing methods used to recover the anisoplanatic PSF is given, and a method used to qualify Taylor's hypothesis, in terms of spatial constraints, concludes the chapter.

Chapter 5 introduces reservoir computing and proposes the adaptation of a machine learning architecture for spatiotemporal prediction. This includes a section on spatiotemporal training and network optimisation. A multi-layer simulation environment for wavefront propagation is described in Chapter 6. Data structures of the resulting time-series ensembles are outlined, and the interface of an optimised artificial neural network architecture is described. Simulation results, employing both CCD read and photon noise to create random data sequences, are detailed in Chapter 7. Lastly, a summary of these results and recommendations for further research are presented in Chapter 8.

In summary, a new method that employs machine learning to improve wide-field astronomical imaging is proposed. This method can be applied to extended, anisoplanatically distorted images, such a shown in Figure 1.2, where the anisoplanatic PSF is estimated using modal wavefront perturbations from multiple reference objects.

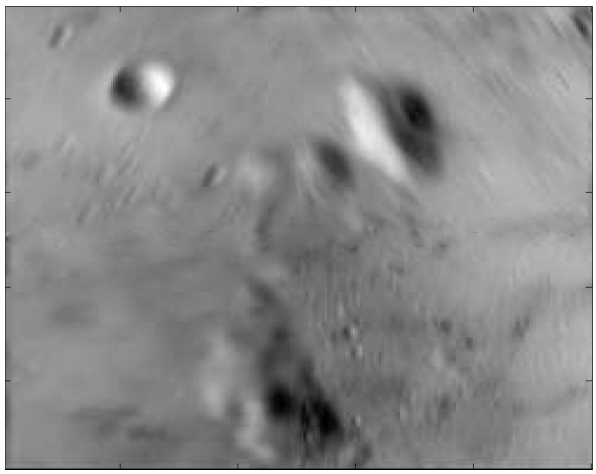

Figure 1.2 Anisoplanatically distorted image of an extended object.

The background, methodology, simulation, and verification of this method are the principal topics contained in the remaining chapters of this thesis. A summary of current restoration techniques, in addition to a proposed interpolation method to minimise boundary distortion resulting in the restoration of images such as Figure 1.2, are outlined in Appendix A. Ex-

perimental methods using laboratory equipment to generate training and verification data used in this study are described in Appendix B. Empirical data ensembles were used for temporal analysis of Zernike coefficients in Chapter 4, and these were included in a discussion on spatial constraints using the Taylor hypothesis. These datasets were acquired during several observation runs at MJUO; details of data acquisitions from three observation runs are listed in Appendix C. Lastly, since the point spread function (PSF) is used extensively throughout this thesis, a derivation of the continuous and discrete PSF is presented in Appendix D.

Chapter 2

Mathematical Preliminaries

This chapter provides the background of mathematical principles and methods used in the development of this work. The underlying proof from which these principles and theorems are based, has been omitted; the objective is to provide a succinct and convenient reference for the mathematical operations used in subsequent chapters.

Section 2.1 defines the notation adopted throughout this thesis, specifically in terms of vectors, matrices and coordinate systems. Since the simulation and experimental results presented in this dissertation are based on measured *a posteriori* time-series data, a background on random variables and processes, subject to statistical methods, is provided in Section 2.2.

A small but fundamental set of linear methods, including regression, singular value decomposition, and the pseudoinverse, are used throughout this work. These methods are applied to data ensembles that represent ill-posed problems, ultimately to find the most appropriate solution. Section 2.3 provides a summary of these linear methods.

Lastly, the application of restorative algorithms is often the principal end-result in astronomical imaging. However, ill-posed problems such as these often require specialised methods to be effective. Tikhonov regularisation is an example of an effective algorithm used in this work for restoration of extended, astronomical images; this algorithm is defined in Section 2.4.

2.1 Notation

2.1.1 Vectors and matrices

A vector is a one dimensional (1D) set of components, such as an array of numbers, represented by a bold, lower-case Roman or Greek letter, e.g., $\boldsymbol{\xi} = \{\xi_1, \xi_2, \cdots, \xi_N\}$. The vector $\mathbf{x}$ is commonly used to abbreviate spatial coordinates within a two dimensional (2D) Cartesian coordinate system, e.g., $\mathbf{x} = \{x, y\}$. Additionally, the spectrum of an image in Fourier space can be represented by $\mathbf{u} = \{u, v\}$. Where relevant in this thesis, a distinction is made between identical letters used for coordinate row vectors, and (typically) column, data vectors.

Lastly, when describing a quantity in three dimensional (3D) Euclidian space, the vector, $\mathbf{r}$, is used, where $\mathbf{r} = \{x, y, z\}$. An alternative to the bold, letter representation for vectors is, $\vec{x}$, however this convention is not used in this dissertation.

Matrices are represented by a capital, bold letter and comprise a J-dimensional (kD) array of numbers. For example, $\mathbf{X}$ is a 2D matrix defined as

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,N} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,N} \\ \cdots & \cdots & \cdots & \cdots \\ x_{M,1} & x_{M,2} & \cdots & x_{M,N} \end{bmatrix}, \tag{2.1}$$

and is composed of elements, $x_{i,j}$, where i and j are row and column indices, respectively; the total number of elements, i.e., the size of $\mathbf{X}$ is given as $M \times N$.

A special matrix used extensively in linear algebra is the identity matrix, $\mathbf{I}$. This is a diagonal matrix with a series of 1's on the diagonal and 0's off the diagonal. It is analogous to unity in numerical algebra and is first used in Section 2.4.

2.1.2 Coordinate spaces

Two coordinate systems are used in this thesis, these are the Cartesian and polar coordinate systems. Since both are used quite differently, each of their applications will be outlined.

As defined in the previous subsection, vectors are used to define coordinates within a J-

dimensional Cartesian coordinate system. In terms of images, a subscript is used to differentiate vectors that are composed of a similar set of components. For example $\mathbf{x}_1$ and $\mathbf{x}_2$ might represent positions in the pupil and image planes of an optical system, where $\mathbf{x}_1 = (x_1, y_1)$, and $\mathbf{x}_2 = (x_2, y_2)$, respectively.

The polar coordinate system is used in this thesis to represent Zernike polynomials defined in Section 3.3, and the angular separation, θ, between two natural guide stars (NGSs). A fixed point, referred to as the *pole*, is defined at the position of observation, such as the location of a telescope on Earth. Two rays, P_1 and P_2, as defined by geometrical optics, can be used to represent the propagation path of planar wavefronts from two NGSs towards the pole. Thus, rays P_1 and P_2 form an angular separation, θ_{Sep}, w.r.t. the pole. In this thesis, an alternative term, *field angle*, is synonymous with angular separation.

The distance from the pole to each NGS is considered infinite, however the distance to separate, thin layers of turbulence is finite. This finite distance is termed, z and the angle between $(P_1 + P_2)/2$ and zenith is θ_z. Considering the angle made between P_1 and P_2 and zenith, when $\theta_{Sep} \ll \theta_z$, the height of a turbulent layer is given as, $h = z\cos(\theta_z)$. The height parameter, h, together with the field angle, θ_{Sep}, are used extensively in simulations detailed in Chapter 6.

Conversions between polar and Cartesian coordinates are performed extensively throughout this work, and can be found in any elementary trigonometry textbook.

2.2 Random variables and random processes

Signals used in most signal processing applications are embedded in an environment of noise and interference. For example, measurements of a signal taken at different times will result in a different outcome - such differences can be in terms of spurious temporal or spatial values. A random or stochastic process is typically used to describe a signal processing problem, where noise and interference are factored into the analysis. In this section, a background to the mathematical analysis used to model real-world application to signal processing problems is presented.

2.2.1 Random variables

A random variable, $\mathscr{X}$, can be defined in terms of the possible number of realisations, x, where $\mathscr{X}(x)$ forms an ensemble of a random process. More specifically, $\mathscr{X}(x)$ is the possi-

ble outcomes of a repeated experiment. Given the statistical tools available, the cumulative distribution function (CDF) describes the probability that a random $\mathscr{X} \leq x$ will occur, and is defined as,

$$F_{\mathscr{X}}(x) = P\{\mathscr{X} \leq x\}. \tag{2.2}$$

A commonly used characterisation of a random variable is defined by its probability density function (pdf), $f_{\mathscr{X}}(x)$ where

$$f_{\mathscr{X}}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{\mathscr{X}}(x). \tag{2.3}$$

Since $F_{\mathscr{X}}(x)$ is monotonic and the summation of all possible outcomes is unity, the pdf only supports natural values.

Random variables are used in this study in an attempt to accurately emulate photon starvation and detector noise, inherent in CCD image sensors. Their respective pdfs define their respective statistical behaviours, given such random events.

Uniform pdf

Given the assumption of a uniform pdf, $\mathscr{X}$ is equally likely anywhere in the interval between x_1 and x_2,

$$f_{\mathscr{X}}(x) = \begin{cases} \frac{1}{x_2 - x_1} & x_1 < x \leq x_2. \\ 0 & \text{elsewhere} \end{cases} \tag{2.4}$$

Gaussian (Normal) pdf

The Gaussian, also referred to as the normal, random variable, $\mathscr{X}$, is defined by the following pdf,

$$f_{\mathscr{X}} = \frac{1}{\sqrt{2\pi}\,\sigma_{\mathscr{X}}} \exp\left\{ -\frac{(x-\mu_{\mathscr{X}})^2}{2\,\sigma_{\mathscr{X}}^2} \right\}, \tag{2.5}$$

where $\mu_{\mathscr{X}}$ is the arithmetic mean, and $\sigma_{\mathscr{X}}^2$ is the variance of the distribution. The Gaussian model is used extensively in a wide range of applications. In terms of this research, the Gaussian model is used in the simulation of CCD read noise.

Justification for using the Gaussian distribution for many problems that require random variables is the central limit theorem. The central limit theorem (CLT) is of great significance in Nature and is used as a basic prior for many of the processes on random data

ensembles used in this research. The CLT states that the average of independent and identically distributed (IID) random variables with a finite variance, will trend towards a normal distribution. If $\mathscr{X}_1, \mathscr{X}_2, \cdots \mathscr{X}_N$ are IID random variables with mean μ and variance σ^2, then the sequence of random variables

$$S_N = \frac{\sum_{i=1}^{N}(\mathscr{X}_i - \mu)}{\sigma\sqrt{N}}, \tag{2.6}$$

will converge in distribution to a standard normal random variable as $N \to \infty$ [26].

Poisson pdf

A Poisson random variable, $\mathscr{X}$, can be described in terms its pdf [12],

$$f_{\mathscr{X}}(x) = \sum_{k=1}^{\infty} \frac{\mu_{\mathscr{X}}^k \exp(-\mu_{\mathscr{X}})}{k!}\delta(x-k), \tag{2.7}$$

where $\mu_{\mathscr{X}}$ is both the mean and variance.

The Poisson distribution is used in this work to simulate the statistical nature of light. For example, if $\mu_{\mathscr{X}}$ is used to represent the expected photon count received by a light detector over many experiments, then the probability of obtaining a photon count is

$$P(x) = f_{\mathscr{X}}(x) = \frac{\mu_{\mathscr{X}}^x e^{-\mu_{\mathscr{X}}}}{x!}. \tag{2.8}$$

Joint probability density functions

Two or more random variables are said to be statistically independent if their joint pdf can be expressed as a product of their individual pdfs. For example, given two random variables, $\mathscr{X}$ and $\mathscr{Y}$, their independence can be stated as

$$f_{\mathscr{X},\mathscr{Y}}(x,y) = f_{\mathscr{X}}(x) f_{\mathscr{Y}}(y). \tag{2.9}$$

If however, two or more random variables are not statistically independent, Bayes theorem can be applied to a joint probability density function (pdf), in the formulation of a conditional pdf. For example, given the same random variables, $\mathscr{X}$ and $\mathscr{Y}$, Bayes' theorem is

used to provide the conditional pdf

$$f_{\mathscr{X}|\mathscr{Y}}(x|y) = \frac{f_{\mathscr{X},\mathscr{Y}}(x,y)}{f_{\mathscr{Y}}(y)}. \tag{2.10}$$

The conditional pdf in Equation 2.10 is continuous. A discrete form of the joint pdf is given as,

$$p_{\mathscr{X}|\mathscr{Y}}(x_i|y_i) = \frac{p_{\mathscr{X},\mathscr{Y}}(x_i,y_i)}{p_{\mathscr{Y}}(y_i)}, \tag{2.11}$$

where i is an integer such that, $0 \geq i \geq N-1$.

2.2.2 Random signals and random processes

Where a random variable assigns values to a set of possible outcomes, a random signal is a sequence of random variables. A set, or ensemble of random signals, each of which constitutes a function of one or more dependent variables, is a random process. For example, an ensemble of time-series functions, each composed of random signals such as Zernike coefficients, is a random process.

A random process is said to be *stationary* if the density functions describing the process are invariant, given a translation of time [14]. Consider a set of random variables, $\mathscr{X} = \{\mathscr{X}(t_1), \mathscr{X}(t_2), \cdots \mathscr{X}(t_N)\}$, and a translated set, $\mathscr{X}' = \{\mathscr{X}(t_1+\tau), \mathscr{X}(t_2+\tau), \cdots \mathscr{X}(t_N+\tau)\}$. For example, if the respective density function, $f_{\mathscr{X}(t_1)}, f_{\mathscr{X}(t_1),(t_2)}, \cdots, f_{\mathscr{X}(t_1),(t_2),\cdots,(t_N)}$ describing $\mathscr{X}$, are identical in form to those describing $\mathscr{X}'$, the process is said to be stationary.

Strict stationarity, as distinct from *wide-sense stationarity*, is a more restrictive requirement. For a random process to be *strictly stationary*, all the high-order pdfs should be invariant under a time translation. Since it is often difficult to verify this, a less demanding form of stationarity can be assumed. Thus, a wide-sense stationary process can be assumed, if $E\{\mathscr{X}(t_1)\}$ is independent of t_1 and $E\{\mathscr{X}(t_1)\mathscr{X}(t_2)\}$ is dependent only on the time difference, $t_2 - t_1$ [14].

In terms of the random signals relating to the effects of atmospheric turbulence, the random processes are assumed to be wide-sense stationary [117]. The most restrictive form of random signals that exist that referred to as *ergodic*. A stochastic process is said to be ergodic in the mean, if its time and average ensembles are equal.

2.2.3 Moments

Consider the random variable, $\mathscr{X}$, for which the statistical average is required. This can be defined by

$$\mu_{\mathscr{X}} = E\{\mathscr{X}\} = \int_{-\infty}^{\infty} x f_{\mathscr{X}}(x)\,\mathrm{d}x, \tag{2.12}$$

where $E\{\cdot\}$ is the expectation operator and $f_{\mathscr{X}}(x)$ is the pdf of $\mathscr{X}$ for values, x. The expectation operator extracts from the random variable the value which the variable is most likely to be, the mean $\mu_{\mathscr{X}}$.

The k^{th} moment of random variable $\mathscr{X}$ is defined as,

$$\mu_{\mathscr{X}}^k = E\{\mathscr{X}^k) = \{\mathscr{X}^k\} = \int_{-\infty}^{\infty} x^k f_{\mathscr{X}}(x)\mathrm{d}x. \tag{2.13}$$

The second, central moment, $\mathscr{X}^2$, where the distribution is centered about the mean $\mu_{\mathscr{X}}$, is the variance, $\sigma_{\mathscr{X}}^2$, defined as

$$\sigma_{\mathscr{X}}^2 = E\{(\mathscr{X} - \mu_{\mathscr{X}})^2\} = \int_{-\infty}^{\infty} (x-\mu_{\mathscr{X}})^2 f_{\mathscr{X}}(x)\,\mathrm{d}x, \tag{2.14}$$

and $\sqrt{\sigma_{\mathscr{X}}^2}$ is the standard deviation.

Considering the case of two or more random variables, such as $\mathscr{X}$ and $\mathscr{Y}$, the expected value of their product is termed the correlation of $\mathscr{X}$ and $\mathscr{Y}$, and is given by

$$R_{\mathscr{X}\mathscr{Y}} = E\{\mathscr{X}\mathscr{Y}\} = \int_{-\infty}^{\infty} xy f_{\mathscr{X}\mathscr{Y}}(x,y)\mathrm{d}x\mathrm{d}y. \tag{2.15}$$

When the expectation is taken of a random variable multiplied with itself, this is known as autocorrelation, $R_{\mathscr{X}\mathscr{X}}$, stated as

$$R_{\mathscr{X}\mathscr{X}} = E\{\mathscr{X}\mathscr{X}\}. \tag{2.16}$$

If the central moments of two random variables, $\mathscr{X}$ and $\mathscr{Y}$, are also considered, the result-

ing operation is termed the covariance, $C_{\mathscr{X}\mathscr{Y}}$, where

$$\begin{aligned} C_{\mathscr{X}\mathscr{Y}} &= E\{(\mathscr{X}-\mu_{\mathscr{X}})(\mathscr{Y}-\mu_{\mathscr{Y}})\} \\ &= k\int_{-\infty}^{\infty}(x-\mu_{\mathscr{X}})(y-\mu_{\mathscr{Y}})f_{\mathscr{X}\mathscr{Y}}(x,y)\mathrm{d}x\mathrm{d}y. \end{aligned} \tag{2.17}$$

The corresponding term used for the expectation of a random variable multiplied with itself, considering the mean of the random variable, is autocovariance,

$$C_{\mathscr{X}\mathscr{X}} = E\{(\mathscr{X}-\mu_{\mathscr{X}})(\mathscr{X}-\mu_{\mathscr{X}})\}. \tag{2.18}$$

2.2.4 Performance metrics

Signal-to-noise Ratio

A common metric used to evaluate the proportion of noise contained within an image is the signal-to-noise ratio (SNR). The SNR metric is often used in signal and image processing to express the quality of data, representing either an Jdimensional signal or image, s, and additive noise, n. The influence of noise, typically modelled on Gaussian statistics, on the signal or image, is defined in terms of the ratio,

$$\mathrm{SNR} = 20\log_{10}\left(\frac{\sigma_s}{\sigma_n}\right)\ \mathrm{dB}, \tag{2.19}$$

where σ_s and σ_n represent the standard deviation of a signal or image, and noise, respectively. Read noise is a major contributor to the overall noise model. The effect of read noise from charge coupled devices (CCD) is discussed in Section 7.5. A brief description of readout noise is given in the following section.

Mean squared error

The mean squared error (MSE), $\varepsilon^2(\mathbf{x})$, is calculated for two ensembles, i.e., an original data series $f(\mathbf{x})$, and an estimate of the original data series, $\hat{f}(\mathbf{x})$. The MSE metric is given by

$$\varepsilon^2(\mathbf{x}) = \frac{1}{N}\sum_{\mathbf{x}\in R}\left[f(\mathbf{x})-\hat{f}(\mathbf{x})\right]^2, \tag{2.20}$$

where N is the size of the ensemble, and $\mathbf{x}$ is a J-dimensional coordinate vector as defined in Section 2.1.

In terms of this research, 1D data are used for time series prediction in Chapter 7, 2D data are required for MSE image comparisons presented in Appendix A.

Normalised mean squared error

The normalised mean squared error (NMSE) metric, $\varepsilon^2_{\text{Norm}}(\mathbf{x})$, is similar to the MSE, however the results of each iteration are normalised with the original data. The NMSE metric is defined as

$$\varepsilon^2_{\text{Norm}}(\mathbf{x}) = \frac{1}{N}\sum_{\mathbf{x}\in R}\left[\frac{f(\mathbf{x})-\hat{f}(\mathbf{x})}{f(\mathbf{x})}\right]^2, \tag{2.21}$$

where each of the terms are identical to those defined for Equation 2.20.

When applying the MSE and NMSE, minimal differences between the two metrics were recorded. This is most likely due to normalisation of input data prior to network prediction.

Similarity metric

The similarity metric was developed as a perceived measure of quality, based on structural similarity [139]. The similarity measure is defined as a function $f(\cdot)$ of three components

$$S(\mathbf{x}_1,\mathbf{x}_2) = f\big[l(\mathbf{x}_1,\mathbf{x}_2), c(\mathbf{x}_1,\mathbf{x}_2), s(\mathbf{x}_1,\mathbf{x}_2)\big], \tag{2.22}$$

where $l(\mathbf{x}_1,\mathbf{x}_2)$ is the luminance of each signal, $c(\mathbf{x}_1,\mathbf{x}_2)$ is a contrast comparison, and $s(\mathbf{x}_x,\mathbf{x}_2)$ is a structure comparison of (typically) 2D matrices $\mathbf{x}_1$ and $\mathbf{x}_2$.

Since all three parameters can be adjusted independently, $f(\cdot)$ in Equation 2.22 leads to a proportional representation between each component [139]

$$\text{SSIM}(\mathbf{x}_1,\mathbf{x}_2) = \big[l(\mathbf{x}_1,\mathbf{x}_2)\big]^{\alpha}\cdot\big[c(\mathbf{x}_1,\mathbf{x}_2)\big]^{\beta}\cdot\big[s(\mathbf{x}_1,\mathbf{x}_2)\big]^{\gamma}, \tag{2.23}$$

where $\alpha>0$, $\beta>0$, and $\gamma>0$ can be adjusted independently, relative to the importance of each component.

The similarity metric $\text{SSIM}(\cdot)$ is used in this thesis to evaluate the quality of a restored image with that of the original. This metric is particularly useful when applied to restored

images of extended objects, and an example is given in Appendix A.

2.2.5 Noise

This subsection provides an analysis of noise sources used in this thesis. Simulations employing each noise condition are presented in Sections 7.5 and 7.6.

Photon Noise

Photon noise sets a limit on our ability to record details of objects, given a photon count imaged on a 2D detector, $P_{x,y}$, over time period, t. The photon flux is given by [11]

$$I_{x,y} = \frac{P_{x,y}}{t}. \tag{2.24}$$

Due to the nature of light, the arrival of photons can be described by the Poisson distribution given by Equation 2.8. As a result, $I_{x,y}$ will be different for each value of (x,y) over the image plane. A measure of the average photon count, $\bar{P_{x,y}}$ can be used to determine the standard deviation, σ_P, given a Poisson distribution, of

$$\sigma_P = \sqrt{\bar{P}_{x,y}}. \tag{2.25}$$

The variation in the flux is referred to as photon noise. Broadly speaking, photon noise is inversely proportional to photon flux. Examples of the effects of photon noise, and subsequent analysis in predicting the effects of turbulence on faint, target objects, is discussed in Section 7.6.

Read Noise

Read noise is additive and is generated from charge to voltage conversion amplifiers on charge coupled devices (CCDs). The measure of readout noise is electrons per pixel and is proportional to both temperature and CCD gain. The distribution used to model readout noise is zero mean Gaussian.

A 2D image can be represented as a 1D noise model [3]

$$g(x) = \sum_{j=1}^{J} \delta(x - x_j) + \sum_{k=1}^{K} \eta_k \, \delta(x - x_k), \tag{2.26}$$

where x_j represents the location of the jth photon event of a light detector, and J is the total number of photon events forming a 2D image represented as a 1D vector. The image is composed of K pixels and η_k is used to denote the random variable corresponding to the read noise at the kth pixel location.

Background Noise

Sky background noise at visual and near infrared wavelengths is a considerable factor when acquiring observational data. This is due to the proportionality of the signal, in terms of photons from a source beacon, to noise, in terms of increased background radiation over a wide FOV [53], [11]. This adverse condition is somewhat countered by the use of short exposure periods, and is constrained by the sampling rate of the WFS, f_S. A sufficiently high sampling rate is required to minimise the temporal decorrelation of the atmosphere. This requirement is dependent on the Greenwood frequency, g_G, defined by Equation 4.11.

2.3 Linear methods

As data are acquired from sensors, measurement errors due to system noise cause inconsistencies when attempting to find a solution. If the system under study is linear, a series of well-known mathematical methods can be employed that provide general solutions to the type of problems discussed in this research. These problems include data acquisition from wavefront sensors, weight minimisation in artificial neural networks (ANNs), and deconvolution in image restoration.

Many of the problems detailed in this study are based on linear methods. However, in the majority of cases the problem is ill-posed. A problem is considered ill-posed if any of the following conditions is satisfied [50]:

1. A nonexistent input vector results in an output vector, i.e., the existence condition is violated.

2. Input data are insufficient to uniquely reconstruct the non-linear dynamics in the solution space.

3. The presence of noise or errors caused by imprecise measurements introduces discontinuity in the reconstruction.

A solution to ill-posed problems can often be found. The issue regarding non-linearity, outlined above in Item 2, is addressed in Subsection 5.3.4. However, the mathematical tools that are used in this research to derive approximate solutions concerning Items 1 and 3 are discussed in this section.

2.3.1 Linear regression

Linear regression, and the Moore-Penrose pseudoinverse extension, are important methods used in virtually every field of science and technology. In terms of this research, both linear regression and the pseudo-inverse are used extensively. For example, K measurements from samples are represented in terms of matrix $\mathbf{A}$, where each measurement represents a row in $\mathbf{A}$, and each column corresponds to an independent, random variable. The rank of matrix $\mathbf{A}$ is defined as the maximum number of independent columns of $\mathbf{A}$.

A dependent random variable, $\mathbf{y}$, is measured, and the set of coefficients that establish a relationship between $\mathbf{A}$ and $\mathbf{y}$ is defined as $\mathbf{x}$. Formally, linear regression attempts to establish a mathematical relationship between one variable, referred to as the dependent variable, and one or more others, known as independent variables.

In vector notation, we can write a set of equations as

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} \tag{2.27}$$

and solve for $\mathbf{x}$. If $\mathbf{A}$ is of full rank and $\mathbf{A}$ has an inverse, the solution is straightforward, $\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{y}$. An example of applying this method is the inverse filter used for image reconstruction from image data and the point spread function, represented in the Fourier domain. However, for over determined and under determined cases, i.e., when $\mathbf{A}$ is not a square matrix, special consideration is required.

When K is greater than the number of unknowns, a least squares (LS) solution can be applied to this problem, where $\mathbf{x}$ is chosen to provide the minimum, average error, applicable to K linear equations. Thus, an error term, $\mathbf{e}$, is used to approximate the best fit of these data,

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{e}. \tag{2.28}$$

The best approximate solution of $\mathbf{x}$ is found by minimising the sum of squares, i.e. $\|\mathbf{y} -$

$\mathbf{A}\mathbf{x}\|^2 = \|\mathbf{e}\|^2$. Without derivation, the least squares solution is

$$\mathbf{x} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\mathbf{A}^{-1}\mathbf{y}. \tag{2.29}$$

Linear regression is equivalent to an artificial neural network (ANN) that does not support a hidden layer, and provides a linear activation function [87]. For complex, J-dimensional input data, ANNs support hidden layers to facilitate the transfer of data to higher dimensional space; regression can then be used to separate clusters of data, i.e., to draw a straight-line through data points that would not have been possible if working in lower dimensional space. This process is is used to classify data and is discussed further in Chapter 5.

2.3.2 Singular value decomposition

A reduction in dimension is often required, especially when a problem is over-determined. Principal component analysis (PCA) is often used to achieve this, however, a covariance matrix from the data is firstly required, from which the eignvalues are obtained. Singular value decomposition (SVD) achieves the same objectives, i.e., determining the eigenvalues and corresponding eigenvectors, without firstly obtaining a covariant matrix.

If $\mathbf{A}$ is a rectangular matrix of dimensions $M \times N$, of rank p, then a $p \times p$ diagonal matrix exists, $\mathbf{S}$, and two isometric matrices $\mathbf{U}$ and $\mathbf{V}$, such that,

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}}, \tag{2.30}$$

where $\mathbf{V}^{\mathrm{T}}$ is the transpose of matrix, $\mathbf{V}$, the column vectors of $\mathbf{U}$ are the eigenvectors of $\mathbf{A}\mathbf{A}^{\mathrm{T}}$, and the columns of $\mathbf{V}$ are the eigenvectors of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$. The singular values on the diagonal of $\mathbf{S}$ are the square roots of the nonzero eigenvalues of both $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ [124].

2.3.3 Pseudoinverse

The requirements for an LS solution to exist are that the columns of $\mathbf{A}$ must be linearly independent and therefore, $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is invertible. A general solution for selecting the best $\mathbf{x}$ for every linear system is to apply the Moore-Penrose pseudoinverse [124]. Essentially, the pseudoinverse technique inverts $\mathbf{A}$, where $\mathbf{A}$ is invertible.

Essentially, the pseudoinverse solution, $\mathbf{x}^{+}$, is found from the pseudoinverse of $\mathbf{A}$, denoted as $\mathbf{A}^{+}$, where the latter is derived from the SVD of $\mathbf{A}$.

If the SVD of $\mathbf{A}$ is $\mathbf{USV}^{\mathrm{T}}$, then the pseudoinverse of $\mathbf{A}$ is,

$$\mathbf{A}^{\mathrm{T}} = \mathbf{VS}^{+}\mathbf{U}^{\mathrm{T}}, \tag{2.31}$$

where the singular values, $\sigma_1, \cdots \sigma_K$ are on the diagonal of $\mathbf{S}$, and the reciprocals, $1/\sigma_1 \cdots 1/\sigma_K$, are on the diagonal of $\mathbf{S}^{+}$. The pseudoinverse of $\mathbf{A}^{+}$ is simply $\mathbf{A}$ [124].

The pseudoinverse method is used in the training of an echo state network, and is applied in Subsection 5.3.2.

2.3.4 Parameter estimation

The estimation of a parameter value is often required from a series of measurements, comprising one or more datasets. In this section, an estimate,$\hat{\Theta}$, of parameter, Θ, from measurements, $\mathbf{x}$, is discussed.

A major application of Bayes inference is maximising the likelihood of an outcome. This can take two forms, the first simply requires evidence of the outcome, the second further supports the evidence with additional priors,

$$P(\mathbf{x}|\mathscr{X}) \approx P(\mathbf{x}|\Theta', \mathscr{X}), \tag{2.32}$$

where

$$\Theta' = \begin{cases} \underset{\Theta}{\arg\max}\ P(\Theta|\mathscr{X}) = \underset{\Theta}{\arg\max}\ P(\mathscr{X}|\Theta)P(\Theta) & \text{MAP} \\ \underset{\Theta}{\arg\max}\ P(\mathscr{X}|\Theta) & \text{ML.} \end{cases} \tag{2.33}$$

Maximum likelihood estimation

Given N measurements of $\mathbf{x}$, where $\mathbf{x} = x_0, x_1, \cdots, x_{N-1}$, a likelihood function, $\Theta\{\cdot\}$, can be constructed, such that the conditional pdf, $f(\cdot)$, is conditional of measurements for a given value of given evidence, Θ. Maximum likelihood (ML), maximises the likelihood function to provide an estimate, $\hat{\Theta}$, and can be expressed as

$$\hat{\Theta}_{ML} = \underset{\Theta}{\arg\max} \left\{ \prod_{i=0}^{N-1} f(x_i|\Theta) \right\}. \tag{2.34}$$

Essentially, ML provides an estimate of the highest probability of an outcome, given the

likelihood of the outcome based on evidence, Θ, according to Bayes rule.

Maximum *a posteriori* estimation

By applying Bayes' rule, *a posteriori* probability is proportional to the product of the likelihood of the data and the *a priori* of the unknowns [32]. Thus, the maximum *a posteriori* (MAP) estimator gives the most likely value of Θ, $\hat{\Theta}_{MAP}$, given the observed data and prior knowledge of the distribution of Θ, $f(\Theta)$.

Within PDFs, partial knowledge can be specified *a priori* and a coarse model can be refined using empirical observations and data. This can be described using Bayesian methods as [56]

$$Posteriori = \frac{\text{Likelihood}}{\text{Evidence}} \times \text{Prior}. \tag{2.35}$$

More formally, Equation 2.35 can be expressed as

$$\begin{aligned}
\hat{\Theta}_{MAP} &= \overset{\arg\max}{\Theta} \left\{ \prod_{i=0}^{N-1} f(\Theta|x_i) \right\} \\
&= \overset{\arg\max}{\Theta} \left\{ \frac{f(\Theta)}{\prod_{i=0}^{N-1} f(x_i)} \prod_{i=0}^{N-1} f(x_i|\Theta) \right\} \\
&= \overset{\arg\max}{\Theta} \left\{ f(\Theta) \prod_{i=0}^{N-1} f(x_i|\Theta) \right\},
\end{aligned} \tag{2.36}$$

where the expression $f(\Theta)\prod_{i=0}^{N-1} f(x_i|\Theta)$ is referred to as the posteriori distribution.

Intuitively, the maximum likelihood term given by Equation 2.34 and shown in Equation 2.36, can be said to be *weighted* by the known prior term, $f(\Theta)$. The maximum *a posteriori* is the maximum outcome of this process.

2.4 Regularisation

Regularisation is a method widely used in image processing, particularly to address ill-posed problems. Tikhonov regularisation is described here, however similar methodologies were invented independently and have been used in a variety of contexts. The method is to accept a non-zero residual term from Equation 2.27, thus returning a smaller solution norm.

This can be expressed as

$$\arg\min\left\{ \|\mathbf{y} - \mathbf{A}\,\mathbf{x}\|^2 + r^2\|\mathbf{y}\|^2 \right\}, \tag{2.37}$$

where the regularisation parameter, r^2, is used to control the weight given to the minimisation of the solution norm, $\|\mathbf{y}\|$, with respect to minimisation of the residual norm, $\|\mathbf{y} - \mathbf{A}\,\mathbf{x}\|^2$ [46].

Since Equation 2.37 is not suitable for numerical computation, the normal regularised solution is given as [46]

$$\mathbf{x} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + r^2\mathbf{I}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{y}, \tag{2.38}$$

where $\mathbf{I}$ is the identity matrix.

Regularisation is extensively used in image processing for restoration. A model based on the Tikhonov regularisation algorithm is discussed by Barakat *et al.* [9]. The restoration of astronomical images using regularisation is outlined in Appendix A.

Chapter 3

Background

This chapter provides background material in two important fields that are highly relevant to this study, namely, optics and imaging. The combination of these topics provide a foundation for a focused discussion on adaptive optics in Chapter 4, and are used throughout the remainder of this dissertation.

There are two ways to describe the path of an optical wavefront. Firstly, geometric optics is used to describe the path an electromagnetic ray, such as broadband light, as it passes through an optical system. Secondly, Fourier optics is used to describe the physical effects arising from the wave nature of light. An example of this is diffraction, where the deviation of light rays from a rectilinear path is defined. Since both fields of optics are used to describe processes used in this dissertation, a brief summary of both will be given in this section. In addition, since modal expansions are used to define wavefront aberrations, details are provided on their definition. In particular, a commonly used orthogonal basis function set is used to describe aberrations over a circular pupil, i.e., Zernike polynomials.

Fundamental imaging concepts are combined with mathematical tools to introduce anisoplanatic image restoration. Spatial and optical transfer functions are defined, in addition to coherent and incoherent imaging, to provide a basis for defining a spatiotemporal image model used in this study. Relating this spatiotemporal model, in terms of a specific modal expansion set and temporal considerations, is detailed. The chapter concludes with a brief discussion on spatial variance and constraints, such as quantisation, that are addressed in subsequent chapters.

3.1 Geometrical optics

Geometrical optics is a branch of optics that uses geometrical relationships, known as the laws of geometrical optics [84], to simplify the manipulation of wavefronts through the interaction of reflective and/or refractive bodies. Thus, the direction of light rays can be traced as they are altered by optical components, such as lenses, mirrors, or prisms. Fundamental to the study of geometrical optics is that light travels in straight lines; such lines are referred to as *rays*, and incorporate an arrow that indicates the direction of propagation.

3.1.1 Optical path lengths and differences

The optical path length (OPL) can be described as the product of the geometric length of a path of light, L, composed of a number of wavelengths that fit the OPL, and the index of refraction, n, of the medium through which it propagates [36],

$$\text{OPL} = L \times n. \tag{3.1}$$

Given that the refractive index of air is unity, and differences in the refractive index of air due to turbulence, n, are typically small and measured in terms of parts per million, the refractivity of air is typically expressed in terms of $N = (n-1) \times 10^6$.

Equation 3.1 states that small fluctuations in the index of refraction, n, results in variations in the OPL. As s result, the OPL is an important parameter when considering the optical path difference (OPD) between two paths of light.

Consider for example the propagation of two rays, R and R', emanating from a point source, S, as shown in Figure 3.1. If both R and R' propagate through a homogeneous medium, i.e., of identical refractive index, then the optical system shown in Figure 3.1 would ensure that both rays were of identical length and converge to focal point f on image plane x. Given an infinite number of rays, the resulting arc r forms a reference, where all points on r are of equal distance to focal point, f.

If, however, both rays passed though an inhomogeneous medium, where each propagation path supported a different refractive index, then, according to Equation 3.1, the optical path of R' would differ from that of R. Consequently, for an infinite number of rays, arc r' will differ to that of reference arc r. Thus, an optical path difference results and this is shown as ΔW(x, y) in Figure 3.1.

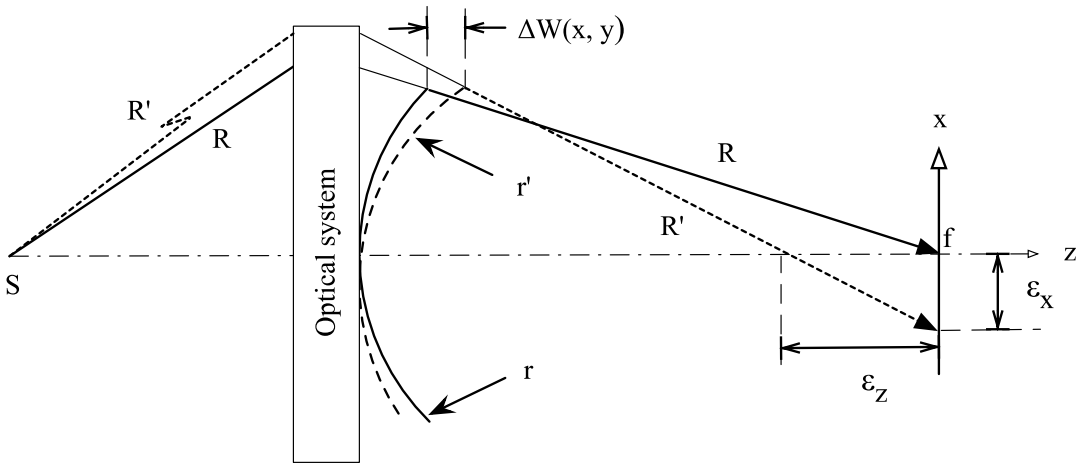

Figure 3.1 Geometric representation showing the relationship between the optical path length (OPL) and the optical path difference (OPD). Adapted from Wyant and Creath [148].

Given this example, the resulting wavefront, r' is said to be *aberrated*. The result of this aberration is the traverse and longitudinal errors shown as ε_x and ε_z in Figure 3.1.

3.1.2 The optical wavefront

A wavefront, generated by a point source at an astronomical distance from an observer and propagating through free space, can be assumed to be effectively planar. Furthermore, if a planar wavefront is not perturbed by an intervening medium, such as the Earth's atmosphere, a point source object when viewed through a perfect imaging system will appear in the image plane as an undistorted, spatially invariant, point spread function. For example, the planar wavefront, $W'(x_0, y_0)$ shown to the left in Figure 3.2, is an example of an unaberrated wavefront just described. If not perturbed by an intervening medium, $W'(x_0, y_0)$ would propagate through a perfect optical system as a spherical wavefront, i.e., reference $S(x, y)$. The result of propagating a point source over a large distance in the absence of any perturbation, would be an Airy disk in the image plane, $I(x, y)$. In terms of the spherical wavefront, the optical path length (OPL), as described in the preceding section, for each ray would be the same.

However, air turbulence adversely affects the shape of a spherical wavefront,as shown. As the wavefront propagates through a turbulent region, shown in Figure 3.2 as Λ, the phase of the wavefront is altered, thus resulting in a wavefront error. Wavefront error, $\varepsilon(x, y)$, is measured with respect to a reference wavefront, and the unit of measure is radians (rad).

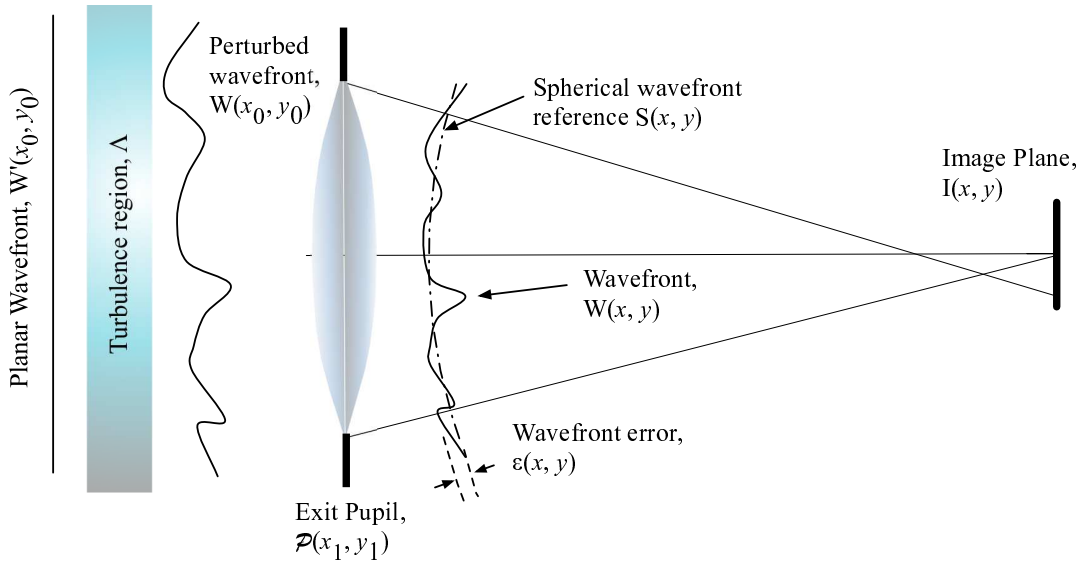

Figure 3.2 Geometric representation of a perturbed wavefront.

3.2 Fourier optics

While geometrical optics is useful for predicting a single point of light on the focal plane of an optical system, it is inadequate to explain the spread of the intensity of light on that plane. For visualisation of the operational aspects and the design of a wavefront sensor, geometric optics is extremely useful, however Fourier optics is required to quantify wavefront aberrations that occur in the pupil and that can be represented by an impulse response function. Thus, Fourier optics deals with the wave nature of light.

Since both near-field and far-field diffraction processes are discussed in this dissertation, both the Fresnel and Fraunhofer approximations will are discussed in this section. Background on diffraction in the generalised pupil is used to introduce the system impulse response function. Following this, the point spread function and optical transfer function, extensively referred to in this study, are introduced.

3.2.1 Diffraction

The diffraction problem, as it is known, is to determine the value of a wavefront or disturbance, U, at some observation point, (x, y), in image space.

In this study an understanding of the two types of diffraction is significant, as this is related

to the formulation of the PSF. In the case of imaging a point-sourced object in the image plane, if the object is focused, the PSF is the Fraunhofer diffraction pattern of the exit pupil. However, if the object is defocused, the PSF is the Fresnel diffraction pattern of the exit pupil [35].

Fresnel diffraction

Fresnel diffraction function is an approximation that describes how a point source of light is distributed in the near-field as light passes through a diffracting aperture plane (x',y'), for the (x,y) observation plane. Approximations, in terms of using quadratic wavefronts to replace spherical wavefronts, result in the Fresnel diffraction integral

$$U(x,y,z) = K_1 \int_{\infty}^{\infty}\int_{\infty}^{\infty} U_{\text{trans}}(x',y') \exp\left\{ j\frac{k}{2z}\left[(x-x')^2(y-y')^2\right]\right\} \mathrm{d}x'\mathrm{d}y', \tag{3.2}$$

where $k = 2\pi/\lambda$, $K_1 = \exp[-j(kz-\omega t)]/j\lambda z$, and $U_{\text{trans}}(\cdot)$ is the transmission function describing the source plane, in terms of amplitude and phase [12].

A Fresnel diffraction calculation is used to simulate the curvature wavefront sensor, where the illumination in two defocused images is required [115]. A simplification of Fresnel diffraction, referred to as the *far-field* condition, is discussed in the next section.

Fraunhofer diffraction

The Fraunhofer far-field approximation applies when the distance, z, from an aperture, D, exceeds $2D^2/\lambda$ [44]. If, however, a converging lens is fitted to the aperture, then the Fraunhofer diffraction pattern can be observed much closer to the aperture, i.e., as the spherical wavefront converges to the focal point, as a result of the lensed system.

The complex field distribution, $U(x,y,z)$, for the far-field is defined by the Fraunhofer approximation,

$$U(x,y,z) = K_2 \int_{\infty}^{\infty}\int_{\infty}^{\infty} U_{\text{trans}}(x',y') \exp\left[-j\frac{2\pi}{\lambda z}(xx'+yy')\right] \mathrm{d}x'\mathrm{d}y', \tag{3.3}$$

and can be described as the Fourier transform of the complex transmission function, $U_{\text{trans}}(\alpha,\beta)$. The complex term, K_2, is defined as [12]

$$K_2 = \frac{\exp[-j(kz-\omega t]}{j\lambda z} \exp\left[\frac{\pi}{\lambda z}(x^2+y^2)\right]. \tag{3.4}$$

The Fraunhofer approximation is valid for most astronomical applications; for example, when applied to 589 nm light passing through a 1-m aperture, the observation distance z must satisfy, $z >> 2600$ km. Clearly, facilitation of an objective, such as the f/13 primary mirror on the 1-m McLellan telescope at MJUO, will result in, $z = 13$ m.

3.2.2 Diffraction in the generalised pupil

To understand how the pupil can be localised and wavefront aberrations measured to extend anisoplanatic imaging, it is instructive to study the effects of diffraction as a consequence of aberrations in the pupil. Diffraction effects are assumed to occur during the propagation of wavefronts from an object to the entrance pupil of an optical system, and then as the wavefront further propagates through an optical system, from the exit pupil to the observation plane.

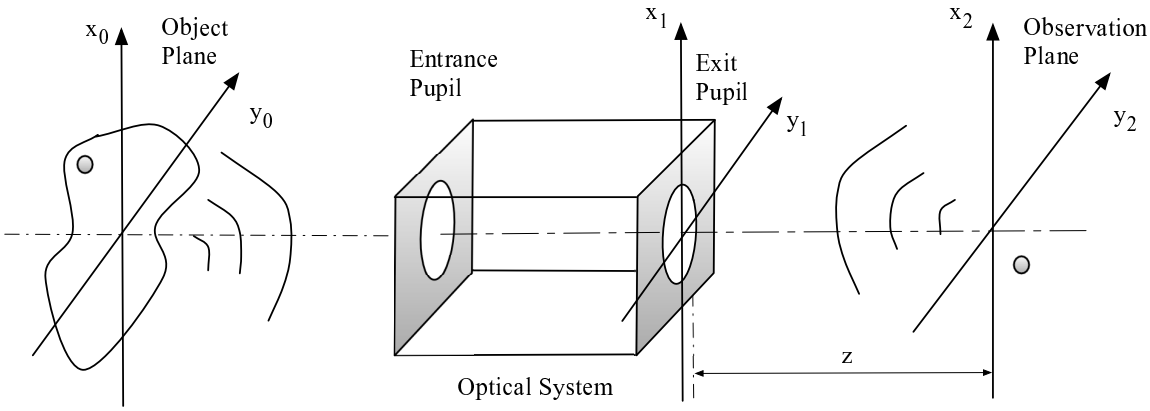

Figure 3.3 Generalised image model.

The generalised pupil function $\mathscr{P}(x,y)$ represents the wavefront in the exit pupil of an optical system [44], and when supported by aperture A, can be defined as

$$\mathscr{P}(x_1,y_1) = \begin{cases} P(x_1,y_1)\exp\left[j\phi(x_1,y_1)\right] & \{x_1,y_1 \in A \,|\, \Re_{x,y}\} \\ 0 & \text{elsewhere,} \end{cases} \tag{3.5}$$

where $P(x,y)$ for a diffraction limited system is simplified to

$$P(x_1,y_1) = \begin{cases} 1 & \{x_1,y_1 \in A \,|\, \Re_{x,y}\} \\ 0 & \text{elsewhere.} \end{cases} \tag{3.6}$$

Wavefront aberrations in the pupil are measured as phase perturbations, $\phi(x_1, y_1)$, and are expressed in terms of the wavefront aberration function, $W(x_1, y_1)$, as

$$\phi(x_1, y_1) = \frac{2\pi}{\lambda} W(x_1, y_1). \tag{3.7}$$

Given the simple optical system shown in Figure 3.3, the image impulse response function, $h(x_2, y_2; x_1, y_1)$, using a single thin lens configuration [12], can be defined as

$$h(x_2, y_2; x_1, y_1) = K \int_{\infty}^{\infty} \int_{\infty}^{\infty} \mathscr{P}(x_1, y_1) \exp\left\{ -j \frac{2\pi}{\lambda z} \left[x_1 x_2 + y_1 y_2 \right] \right\} dx_1 dy_1, \tag{3.8}$$

Equation 3.8 is the Fraunhofer diffraction pattern of the exit pupil, where z is the distance from the exit pupil to the observation plane as shown in Figure 3.3.

It can be shown, with the following change of variables, $u = \frac{x_1}{\lambda z}$, and $v = \frac{y_1}{\lambda z}$, that the RHS of Equation 3.8 is the Fourier transform of the pupil function [12],

$$\begin{aligned} h(x_2, y_2) &= \int_{\infty}^{\infty} \int_{\infty}^{\infty} \mathscr{P}(\lambda z u, \lambda z v) \exp\big\{ -j2\pi(u x_2 + v y_2) \big\} du dv \\ &= \mathrm{FT}\big\{ \mathscr{P}(\lambda z u, \lambda z v) \big\}. \end{aligned} \tag{3.9}$$

Equation 3.9 is used in Subsection 3.4.1, where measurements of the system response function are made in the exit pupil of an optical system, using wavefront sensing. The convolution of $h(\cdot)$ with an object in the object plane is the basis for image formulation.

Within the visual spectrum, ground-based astronomical imaging systems do no to operate at the diffraction limit of their optics. The resulting images of exo-atmospheric objects thus appear distorted. A significant contribution to such distortion is due to phase perturbations, shown as $\phi(\cdot)$ in Equation 3.5, caused by atmospheric turbulence. Thus, the generalised pupil function $\mathscr{P}(\cdot)$ is used to express image formulation as shown by Equation 3.9.

3.3 Modal expansions

Motivated by mathematical convenience, signals are typically represented by the linear combination of a set of basis functions. For example, let a set of basis functions be represented as, $\varphi_{-N}(\xi)\cdots$, $\varphi_{-1}(\xi)$, $\varphi_0(\xi)$, $\varphi_1(\xi)\cdots$, $\varphi_N(\xi)$, where N can approach infinity.

Almost any arbitrary function, $x(\xi)$, can be represented over an interval, (T_a, T_b) by expansion of

$$x(\xi) = \sum_{n=1}^{\infty} a_n \varphi(\xi), \tag{3.10}$$

where the coefficients, a_n, can be complex and are chosen such that successive orders of the basis function set can be applied to achieve greater accuracy. This is often referred to in the literature as the *Gram-Schmidt procedure* [89].

A desired property for a set of basis functions set is to support finality of the coefficients. This property allows any given coefficient to be determined without knowing any other coefficient. To achieve this, basis functions require orthogonality over an interval, i.e, (T_a, T_b), to be termed a complete orthogonal set [42]. The condition for orthogonality of a real basis set over T_a to T_b, and for all m and n, is

$$\int_{T_a}^{T_b} \varphi_n(\xi)\, \varphi_m(\xi)\, \mathrm{d}\xi = \|\varphi_m\|^2\, \delta_{nm} = \begin{cases} 0 & m \neq n \\ \|\varphi_m\|^2 & m = n, \end{cases} \tag{3.11}$$

where $\|\varphi_m\|$ is the norm of the function and δ_{mn} is the Kronecker delta function. If $\|\varphi_m\| = 1$ for all m in Equation 3.11, the basis functions are *orthonormal*.

An important requirement for a signal is that the energy, E, can be represented in terms of coefficients. In the case of a real, orthonormal basis function set,

$$E = \sum_{n=0}^{N} a_n^2 \|\varphi_n\|^2. \tag{3.12}$$

This is one aspect of a general theorem known as *Parseval's theorem*.

3.3.1 Zernike polynomials

Considering the preceding discussion, Equation 3.10 can be used to represent the phase perturbations of an optical wavefront, $\phi(x,y)$, as imaged in a circular aperture,

$$\phi(x,y) = \sum_{i=1}^{\infty} a_i Z(x,y), \tag{3.13}$$

where $Z(x,y)$ is the Zernike polynomial and the decomposition in terms of the Zernike basis functions is referred to as the modal expansion of $\phi(x,y)$ [117].

Zernike polynomials are 2D orthonormal basis functions commonly used to define optical aberrations over a circular region. They represent the statistical eigenfunctions of optical distortions that quantitatively classify each aberration using a set of polynomials.

Although the set of Zernike polynomials is considered sub-optimal, [115] , the use of Zernike modes as a basis set provides acceptable performance [131] and is used in the analysis presented in this thesis.

This set of polynomials is defined as

$$\left.\begin{aligned} Z_{i=even}(\rho,\theta) &= \sqrt{n+1}R_n^m(\rho)\sqrt{2}\cos(m\theta) \\ Z_{i=odd}(\rho,\theta) &= \sqrt{n+1}R_n^m(\rho)\sqrt{2}\sin(m\theta) \end{aligned}\right\} \quad m \neq 0, \tag{3.14}$$

$$Z_i(\rho) = R_n^0(\rho)\sqrt{2},\, m = 0, \tag{3.15}$$

where ρ is the normalised aperture and is defined in Cartesian coordinates as, $\sqrt{x^2+y^2}$, $\theta = \arctan(y/x)$ i is a single index numbering scheme adopted by Noll [98], and the terms m and n are the azimutal and radial order, respectively. The radial polynomial function, $R_n^m(\rho)$, is defined as

$$R_n^m(\rho) = \sum_{s=0}^{(n-m)/2} \frac{(-1)^n\,(n-s)!}{s!\left[\frac{(n+m)}{2}-s\right]!\left[\frac{(n-m)}{2}-s\right]!}\rho^{(n-2s)}. \tag{3.16}$$

Each Zernike mode can be represented by a 2D image, commonly referred to as a *phase map*. For example, an aberrated PSF is represented by K phase maps on the pupil plane.

The linear combination of K aberrations over a unit circle of radius R results in an approximation of the phase perturbation

$$\phi(R\rho,\theta) \approx \sum_{i=2}^{K} a_i Z_i(\rho,\theta), \tag{3.17}$$

where R is the radius of the aperture, and when $K = \infty$, (3.17) is an exact representation of the phase. Additionally, the piston term, Z_1, has been removed in (3.17), as is common for single aperture instruments.

The Zernike coefficients are defined as

$$a_i = \int_0^1 \int_0^{2\pi} \mathscr{W}(\rho)\phi(R\rho,\theta)Z_i(\rho,\theta)\,\mathrm{d}\theta\,\mathrm{d}\rho, \tag{3.18}$$

and where $\mathscr{W}(\rho)$ is the pupil weighting function given as

$$\mathscr{W}(x,y) = \begin{cases} \frac{1}{\pi} & \rho \leq 1 \\ 0 & \rho > 1. \end{cases} \tag{3.19}$$

Equation (3.19) is a modal description of the wave front over the exit pupil of an imaging system. However, Ragazzoni showed that a smaller, circular portion of a wave front on the pupil can be described by another ensemble of Zernike coefficients [109]. This important result is applied in Subsection 4.4.4 by using a modal representation to describe wavefront aberrations over both isoplanatic and anisoplanatic regions for predicting the spatially variant PSF over a wide FOV.

Several examples of Zernike polynomials are shown in Figure 3.4. This thesis makes extensive use of Zernike coefficients however as stated above, Zernike polynomials are suboptimal, in terms of being strictly orthogonal. An example of an uncorrelated, statistically independent set would be the Karhunen-Loêve functions. In terms of the Kolmogorov spectrum, this set of functions cannot be expressed analytically [98], however in practice they can be expressed in terms of Zernike polynomial expansions [115].

3.4 Intensity and spectral transfer functions

In this section two functions are introduced that are used as the basis for the design of optical systems. The intensity (amplitude) spatial distribution of an optical system is first discussed. This is commonly referred to as the point spread function (PSF). Secondly, the spatial frequency characteristics of an optical system are given by the optical transfer function (OTF). Both the PSF and OTF are close analogues to the impulse response function and system transfer function used in digital signal processing. Where the former pair are used for spatial design of optical systems, the latter are generally used as temporal design tools for digital systems.

3.4.1 The point spread function

In optical signal processing, the response of a system to a point source of light is a 2D impulse response function. This function describes the spread of light intensity (irradiance) over an observation plane. This function is commonly referred to as a point spread function (PSF) and in terms of astronomical imaging, represents both the imaging system and the

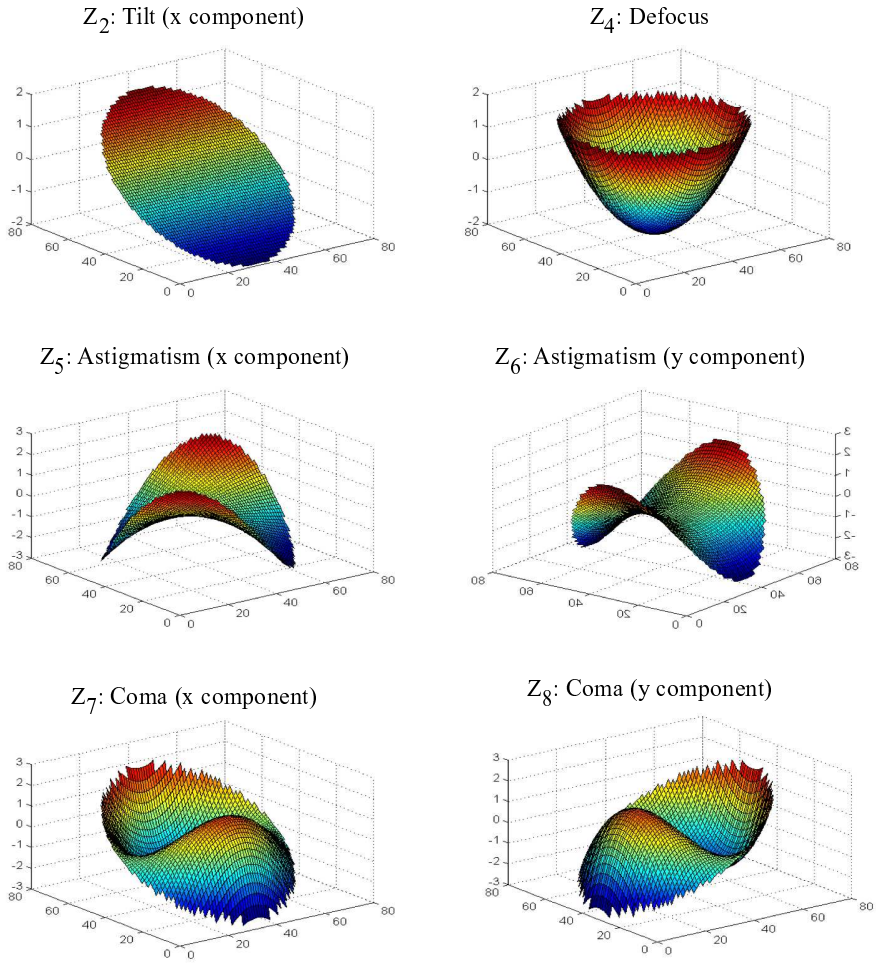

Figure 3.4 Examples of Zernike polynomials with their corresponding Seidel aberrations.

effects of turbulence in the atmosphere.

Mathematically, if H is the system response function and $\delta(\cdot)$ is the delta function, the PSF can be defined as [40]

$$h(x_2, y_2; x_0, y_0) = H\{\delta(x_0 - x_2, y_0 - y_2)\}, \tag{3.20}$$

where (x_0, y_0) are defined as coordinates in object space and (x_2, y_2) are defined as coordinates in image space.

The output of a linear system, $g(x_2, y_2)$, can be defined in terms of an object to be imaged,

$f(x_0, y_0)$, as a sum of weighted point sources [40]

$$\begin{aligned} g(x_2, y_2) &= H\big[f(x_2, y_2)\big] \\ &= H\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_0, y_0)\,\delta(x_2 - x_0, y_2 - y_0)\mathrm{d}x_0\,\mathrm{d}y_0\right]. \end{aligned} \tag{3.21}$$

If H is a linear operator, the additive property can be extended to the integrals

$$g(x_2, y_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H\Big[f(x_0, y_0)\,\delta(x_2 - x_0, y_2 - y_0)\Big]\,\mathrm{d}x_0\,\mathrm{d}y_0. \tag{3.22}$$

Since $f(x_0, y_0)$ is independent of x_2 and y_2, and using the homogeneity property

$$g(x_2, y_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_0, y_0) H\Big[\delta(x_2 - x_0, y_2 - y_0)\Big]\,\mathrm{d}x_0\,\mathrm{d}y_0. \tag{3.23}$$

Substituting Equation 3.20 into Equation 3.23, gives

$$g(x_2, y_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_0, y_0)\, h(x_2, y_2; x_0, y_0)\,\mathrm{d}x_0\,\mathrm{d}y_0. \tag{3.24}$$

Equation 3.24 is called the superposition or Fredholm integral of the first kind. In this form $h(\cdot)$ can be either spatially variant or spatially invariant.

If $h(\cdot)$ is spatially invariant, a single PSF is applied over the image, i.e. for each coordinate in object space, (x_0, y_0), an operation is performed with the object and PSF in image space, (x_2, y_2). Thus, if the spatially invariant point spread function (SIPSF) meets the condition, $h(x_2, y_2; x_0, y_0) = h(x_2 - x_0, y_2 - y_0)$, then Equation 3.21 can be expressed as a convolution [2]

$$\begin{aligned} g(x_2, y_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_0, y_0)\, h(x_2 - x_0, y_2 - y_0)\,\mathrm{d}x_0\,\mathrm{d}y_0. \\ &= f(x_2, y_2) \odot h(x_2, y_2). \end{aligned} \tag{3.25}$$

An example of a SIPSF that supports a consistent degradation over the entire image plane is referred to as *blurring*. Blurring results if an image is taken of an object either not in focus, or if motion has occurred with respect to the image plane during an exposure. In both cases a single SIPSF can be used to describe the resulting degradation.

If however, multiple, inconsistent degradations occur on regionalised sections of the image, as is the case for example with astronomical turbulence, where an analogy of multiple small lenses are placed and moved between the object and images planes [134], then the spatially variant point spread function (SVPSF) is required to fully describe the degradation.

A derivation of the PSF from first principles is provided in Appendix D.

3.4.2 The optical and modulation transfer functions

In addition to the knowing the intensity, also referred to as the amplitude [42], distribution of an optical system, the spatial frequency component is useful in providing a measure of the complex weighting factor applied to the frequency of the optical system, with respect to a weighting factor for the zero-frequency component [44]. The latter is known as the optical transfer function (OTF).

Information concerning the spatial distribution of frequencies, both for short and long exposures in the pupil, is available from the OTF, which is defined as

$$H(u,v) = \mathrm{FT}\big\{h(x_2,y_2)\big\}, \tag{3.26}$$

where $h(x_2,y_2)$ is the PSF, and FT is the Fourier transform operation.

The modulation transfer function (MTF) is taken as the modulus of the OTF, i.e.

$$M(u,v) = |H(u,v)|. \tag{3.27}$$

The optical transfer functions, $H(\cdot)$, for a diffraction limited and aberrated point source, are shown in Figure 3.5 (a) and (c), respectively. The corresponding modulation transfer functions, $M(\cdot)$, are shown in Figure 3.5 (b) and (d).

An increase of wavefront aberrations within an optical system affects the spatial frequencies that can be transmitted. This characteristically results in a narrowing of the OTF and MTF. Correspondingly, the point spread function broadens with increased wavefront aberration as the intensity distribution is spread over a wider spatial region. Such characteristics are evident in the long exposure point spread function discussed in Section 3.6.

It is common practice to normalise the OTF and MTF when describing an optical system.

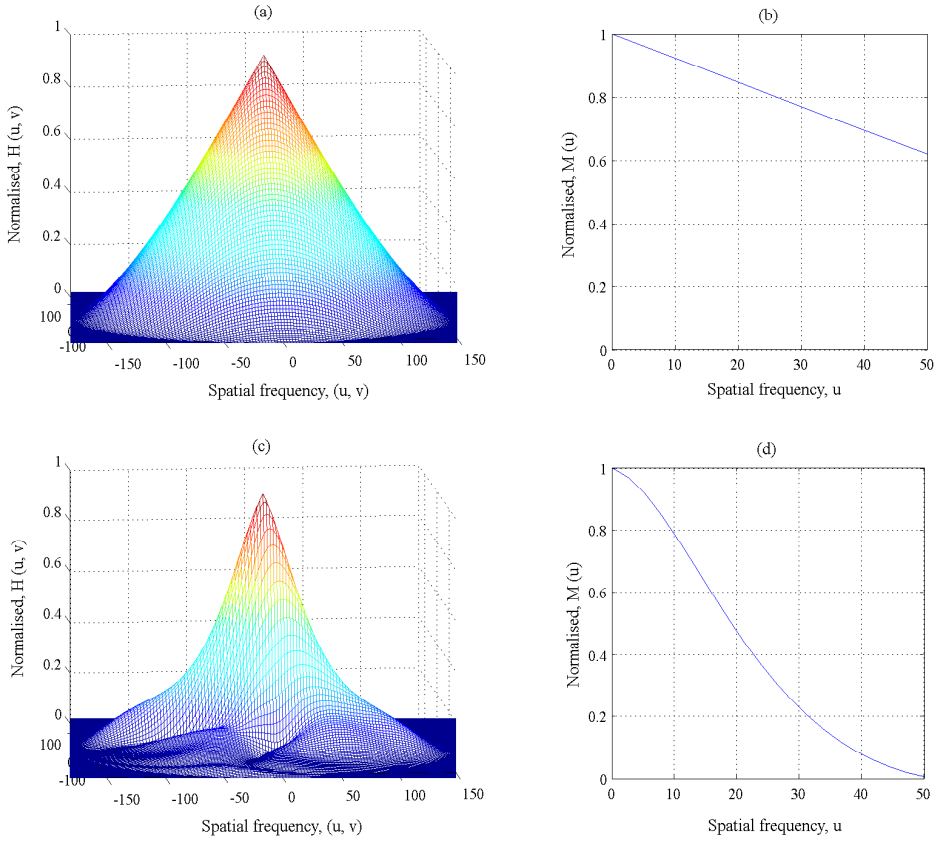

Figure 3.5 Optical and modulation transfer functions: (a) diffraction limited OTF; (b) diffraction limited MTF; (c), aberrated OTF; (d) aberrated MTF.

Such practice is continued in this work. Since the OTF can be used to represent either coherent or incoherent light sources, a more rigourous study of the OTF is the subject of the next section.

3.5 Coherent and incoherent imaging

As will be shown in this section, the incoherent optical transfer function (OTF), represented as $\mathscr{H}(u,v)$, is the normalised autocorrelation of the pupil function, $\mathscr{P}(x_1,y_1)$. Additionally, this is also the Fourier transform of the squared modulus of the PSF. However, in the case of the coherent OTF, $H(u,v)$ this simply results in the rotation of the pupil function. In terms of image intensity, the coherent OTF is considered to be linear. The OTF is analogous to the frequency response of a time-domain filter. The following discussion will attempt to provide some insight into these characteristics.

The coherence of light is an important property that requires careful consideration when designing and using an optical system. Light is never entirely monochromatic nor is it totally coherent [44]. Thus, in practice, partial coherency is considered. However in terms of this discussion, only the extremes of coherency are considered.

A general imaging system function can be expressed in the Fourier domain as

$$G(u,v) = F(u,v)\,H(u,v), \tag{3.28}$$

where $G(u,v)$ is the system output, $F(u,v)$ is the system input, and $H(u,v)$ is the OTF. The system output is a convolution of the PSF and the object in view, $f(x_1,y_1)$; this is equivalent to taking the inverse Fourier transform of Equation 3.28.

As discussed in Section 3.2, the PSF of a generalised imaging system is the Fourier transform of the pupil function. For convenience, Equation 3.9 is repeated as

$$h(x_2,y_2) = \mathrm{FT}\big\{P(\lambda z u, \lambda z v)\big\}. \tag{3.29}$$

The system transfer function using coherent light is found by taking the Fourier transform of Equation 3.29

$$\begin{aligned} H(u,v) &= \mathrm{FT}\big\{h(x_2,y_2)\big\} \\ &= \mathrm{FT}\;\mathrm{FT}\big\{P(\lambda z u, \lambda z v)\big\} \\ &= P(-\lambda z u, -\lambda z v), \end{aligned} \tag{3.30}$$

This last term in Equation 3.32, shows negative parameters for $P(\cdot)$; this result is simply a 180° rotation of the pupil function.

However, for incoherent light, the squared modulus of the PSF is required, i.e. $|h(x,y)|^2$. Thus, the expression for the general imaging system function given in Equation 3.28, can be repeated for the incoherent light as

$$G(u,v) = F(u,v)\,\mathscr{H}(u,v), \tag{3.31}$$

where, now, $\mathscr{H}(u,v)$ is the incoherent transfer function, i.e., the incoherent OTF.

If the process in Equation 3.29 is repeated for the incoherent OTF, we find,

$$\begin{aligned} H(u,v) &= \mathrm{FT}\left\{|h(x_2,y_2)|^2\right\} \\ &= \int\int P(\lambda z u', \lambda z v') P(\lambda z u' - \lambda z u, \lambda z v' - \lambda z v) dv' du' \\ &= P(\lambda z u, \lambda z v) * P(\lambda z u, \lambda z v). \end{aligned} \tag{3.32}$$

As a result of this analysis, Equation 3.32 demonstrates that for an incoherent imaging system, the OTF is simply the autocorrelation function of the pupil. Due to the adverse conditions of imaging through air turbulence, the incoherent OTF is used for astronomical imaging problems. In addition, the generalised pupil function, $\mathscr{P}(x_1, y_1)$, should be employed to incorporate phase aberrations in the pupil [117].

Cross sections of both the incoherent, $h(x_2, 0)$, and coherent, $h_C(x_2, 0)$, PSF are shown in Figure 3.6. The imaging system shown includes a converging lens to insure the amplitude distribution, corresponding to the Fraunhofer diffraction approximation, is achieved if an incoherent light field is received from a point source at infinity. Such an intensity pattern would result, for example, if an imaging system was used to image a natural guide star (NGS) beacon, in the absence of turbulent air; thus, the resolution of the optical system would be working to the diffraction limit of its optics.

However in consideration of a coherent light source, such as produced by a laser, all points on the wavefront have a fixed phase relationship and alternate in unison. As a result, convolution with the optical system, as shown by Equation 3.20, is performed in complex amplitude and is thus linear only in complex amplitude. For incoherent light sources, optical systems are linear in intensity distribution.

3.6 Long and short exposure imaging

The exposure time used to capture a science object, as it appears in the exit pupil of a telescope, has a significant effect on the PSF and OTF. For example, over short time intervals, it has been shown that the atmosphere is effectively *frozen*. The resulting images are referred to as *speckle* patterns and are the subject of a discipline referred to as speckle imaging [117]. In contrast to speckle imaging, long exposures, required to image faint stars, i.e., $10 \leq m_v \leq 16$, broadens the PSF.

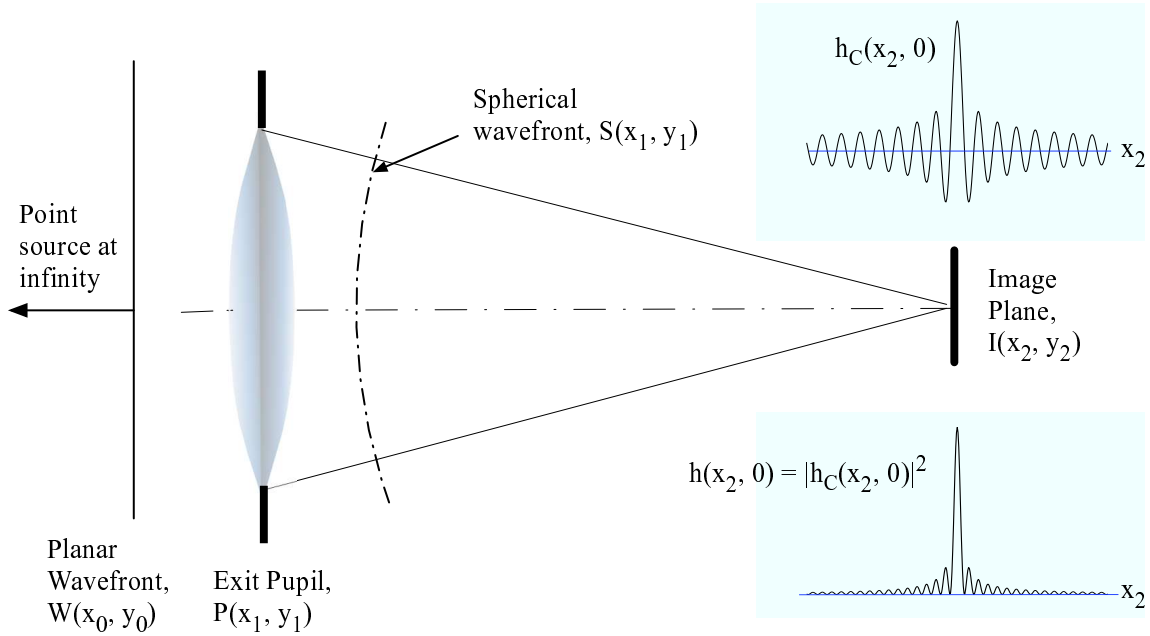

Figure 3.6 Diffraction limited PSF composed from either incoherent, $h(\cdot)$, or coherent, $h_C(\cdot)$, light from a point source at infinity.

It has been estimated that the limit for a short exposure image is less than 20 ms [117], however parameters, such as the altitude of the turbulent layer, diameter of the objective, and Fried's coherence length, also need to be considered [49]. As mentioned, short exposure images have a speckled appearance, however the continuous evolution of tilt aberrations dominate over longer exposure periods, smearing out the speckles. Knowledge of the short and long exposure PSF and OTF functions provide further insight into astronomical imaging, and the amount of compensation required. Thus, two groups of PSFs and associated transfer functions are briefly reviewed in this section, the short and the long exposure PSFs, h_{SE} and h_{LE}, and their respective OTFs, $\mathscr{H}_{SE}$ and $\mathscr{H}_{LE}$.

3.6.1 Long exposure transfer functions

As outlined in the introduction to this section, the long exposure transfer function is predominately affected by tilt aberrations that effectively shift, but do not distort, the short exposure PSF. This has the effect of a low-pass filter that severely attenuates high-order OTF frequencies.

The long exposure PSF is given as an averaged ensemble and is defined in terms of the long

exposure image function

$$\langle g_{\mathrm{LE}}(x_2,y_2)\rangle = f(x_0,y_0) \odot \langle h_{\mathrm{LE}}(x_2,y_2)\rangle + \langle \eta(x_2,y_2)\rangle, \tag{3.33}$$

where the term $\eta(x,y)$ is noise.

The corresponding system function for long exposures is given by

$$\langle G_{\mathrm{LE}}(u,v)\rangle = F(u,v) \cdot \langle \mathscr{H}_{\mathrm{LE}}(u,v)\rangle + \langle N(u,v)\rangle, \tag{3.34}$$

The long exposure, generalised OTF for incoherent light, $\langle \mathscr{H}_{\mathrm{LE}}(u,v)\rangle$, is given by [48]

$$\begin{aligned} \langle \mathscr{H}_{\mathrm{LE}}(u,v)\rangle &= \exp\left[-0.5\,D(\mathbf{u})\right] T(u,v) \\ &= \exp\left[-3.44\left(\frac{\lambda z|\mathbf{u}|}{r_0}\right)^{5/3}\right] T(u,v), \end{aligned} \tag{3.35}$$

where $D(\mathbf{u})$ is the structure function defined later in Equation 4.6, r_0 is the Fried coherence length, $\mathbf{u}$ represents the displacement vector in the aperture, and $T(\cdot)$ is the transfer function of the telescope.

The influence of either the atmospheric or telescope transfer functions in Equation 3.35 has been determined [113]: the former is dominant when aperture $D > r_0$, and the latter when $D < r_0$.

3.6.2 Short exposure transfer functions

As discussed in Section 3.3, wavefront aberrations such as tilt can be represented by Zernike modes Z_2 and Z_3. Such aberrations over a long exposure result in the displacement of the PSF, $h_{\mathrm{LE}}(x_2,y_2)$, in the image plane; the shape of the short exposure PSF is not affected by Zernike terms Z_2 and Z_3. This result is utilised in anisoplanatic image restoration in Appendix A.

In Equation 3.32, the short exposure PSF, $h(x,y)$, is defined as the Fourier transform of the pupil function. The short exposure (SE) OTF is then the Fourier transform of the short exposure PSF, which is also the autocorrelation of the pupil function.

However, additional insight into the dependencies on the telescope parameters, with respect

to the pupil function, can be gained by [117]

$$\mathscr{H}_{SE}(u,v) = \exp\left\{ -3.44\Big(\frac{\lambda z|\mathbf{u}|}{r_0}\Big)^{5/3}\left[1-\alpha\Big(\frac{\lambda z|\mathbf{u}|}{D}\Big)^{1/3}\right]\right\} T(u,v). \tag{3.36}$$

Equation 3.36 is similar to Equation 3.35, with the exception of the term in square brackets. This distinction between short and long exposures places additional emphasis on telescope diameter, D, as a function of the height of the turbulence, z, i.e. $\alpha = 0$ in Equation 3.35 for long exposure images, and $\alpha \neq 0$ for short exposure images. The consequence of this is that for short propagation distances, the SE OTF is only susceptible to phase; longer propagation distances result in contributions, both in terms of phase and amplitude [62].

In this study, the propagation model presented in Section 6.2 is based on short exposure periods over relatively short wavefront propagation distances. This is primarily due to defining parameters to ensure both temporal and spatial sampling criteria are met. During simulations, continuous estimates of the short exposure PSF for a target are provided; the long exposure PSF and OTF functions can be derived from such ensembles, and these results used to provide a basis for modelling and performance comparison.

3.7 The Image Model

Since an object, $f(\cdot)$, can be represented as a set of weighted point sources, the response of a linear system observing that object is the superposition of weighted and shifted versions of the PSF.

Given the coordinate systems defined in Figure 3.3, where (x_2, y_2) is used represent the image plane, and (x_0, y_0) the object plane, the PSF can be written as a function transcending two planes, i.e., $h(x_2, y_2; x_0, y_0)$. If the object being observed is $f(x_0, y_0)$, the image recorded can be written

$$g(x_2, y_2) = \int\int f(x_0, y_0) h(x_2, y_2; x_0, y_0) \mathrm{d}x_0 \mathrm{d}y_0 + \eta(x_2, y_2), \tag{3.37}$$

where $\eta(x_2, y_2)$ represents the inevitable presence of noise in the process. This represents a continuous model for the imaging process.

Without loss of generality and recognising that the image sensors, such as charge coupled devices (CCDs), are inevitably discrete in nature, a discrete equivalent to (3.37) can be

written

$$g(p,q) = \sum_{k=1}^{N}\sum_{l=1}^{M} f(k,l)h(p,q;k,l) + \eta(p,q), \tag{3.38}$$

where (p,q) and (k,l) are the indices of discrete coordinates in image and object spaces, respectively. The noise sources $\eta(p,q)$ comprise CCD sensor read noise and photon noise. These noise sources are modelled using Gaussian and Poisson distributions [26], respectively.

In general, $h(p,q;k,l)$ is dependent both on (p,q) and (k,l), i.e. variations of the PSF depend on the location of the source point. In a significant number of imaging situations, however, a spatially invariant model can be adopted, such that (3.38) becomes a convolution, thus

$$g(p,q) = \sum_{k=1}^{N}\sum_{l=1}^{M} f(k,l)h(p-k;q-l) + \eta(p,q). \tag{3.39}$$

Atmospheric turbulence alters the phase of wavefronts that can be assumed, planar. Phase alterations over the pupil plane result in distortions at the image plane; such distortions are modelled by the deformation of the PSF.

As discussed in Section 3.5, the PSF for astronomical imaging problems is

$$h(x_2,y_2) = \mathrm{FT}\big\{\mathscr{P}(\lambda z u, \lambda z v)\big\}. \tag{3.40}$$

When also considering the area of exit pupil, A_p, the relationship between the wave front and PSF is given as [42],

$$h(x_2,y_2) = \frac{A_p}{\lambda^2 d^2}\left\|\mathrm{FT}\Big\{P(x_1,y_1)\exp[-j\frac{2\pi}{\lambda}W(x_1,y_1)]\Big\}\right\|^2, \tag{3.41}$$

where FT is the Fourier transform operator, d is the distance from the exit pupil to the image, $P(x_1,y_1)$ is the exit pupil function defined in Equation 3.6, λ is the wavelength, and $W(x_1,y_1)$ is the wavefront distortion defined by Equation 3.7 in Section 3.2.

Essentially, Equation (3.41) relates the PSF to the wavefront $W(\cdot)$, as the Fourier transform of the generalised pupil function, $\mathscr{P}(\cdot)$, as defined by Equation 3.5. However, there are two groups of wavefront distortions that affect the PSF differently. The first group comprises low-order aberrations, commonly referred to in the literature as tilt. Such distortions, i.e., Zernike modes Z_2 and Z_3, result in the displacement, rather than distortion of the PSF. The

second group is composed of higher order aberrations, i.e., $Z_2, Z_3, \cdots, Z_\infty$; such wavefront distortions result in the deformation of the PSF.

Most image applications treat the PSF as spatially invariant, however since it has been shown in the previous section that anisoplanatic imaging requires determination of the SVPSF, this study makes extensive use of this function. Figure 3.7 (a) and (c) shows the image and y-cross-section of an unaberrated PSF; sub-figures (b) and (d) show an aberrated PSF, respectively.

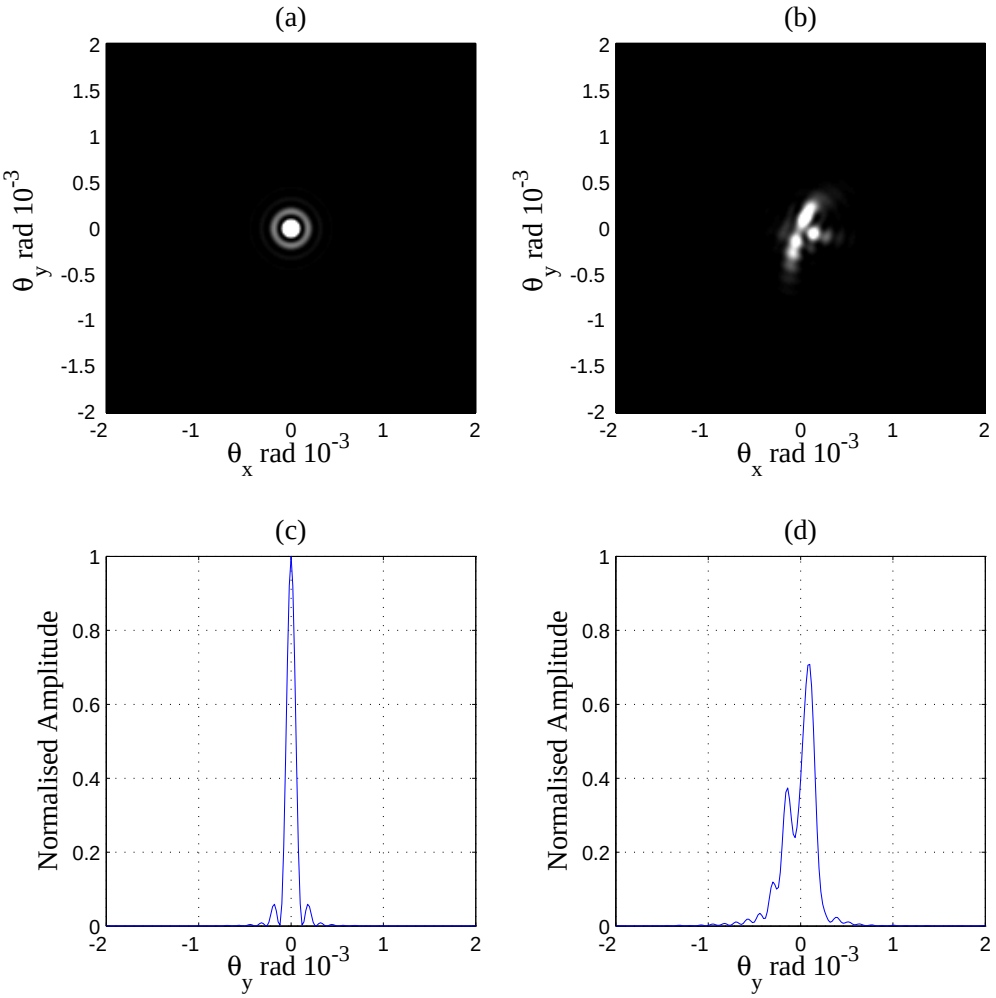

Figure 3.7 Point spread functions: (a) & (c) unaberrated PSF; (b) & (d), aberrated PSF; 2D image and y-axis cross-section, respectively.

The set of Zernike coefficients used to generate the unaberrated and aberrated PSFs, comprising the first and second rows in Table 3.1, is shown in Figure 3.7, respectively. The mean squared error metric (MSE) was used to calculate each Zernike coefficient, using

Table 3.1 Comparative study of PSF degradation using Zernike coefficients (radians).

Comparative degradation	$\mathbf{a_4}$	$\mathbf{a_5}$	$\mathbf{a_6}$	$\mathbf{a_7}$	$\mathbf{a_8}$	$\mathbf{a_9}$	$\mathbf{a_9}$
$r_0 \geq D$	$\approxeq 0$	$\approxeq 0$	$\approxeq 0$	$\approxeq 0$	$\approxeq 0$	$\approxeq 0$	$\approxeq 0$
$r_0 < D$	-0.1026	0.0358	0.0541	-0.0003	0.0001	$< 10^{-4}$	$< 10^{-4}$
$r_0 \ll D$	-4.2673	-1.3828	3.0889	-0.2739	2.6356	-1.1830	1.7106

a time-series ensemble of 500 samples. The last row in Table 3.1 lists values for severe turbulence profiles,where Fried's coherence length, r_0, is significantly less than the size of the aperture, D. Each ensemble was generated using wavefront propagation simulations discussed in Section 6.2.

In summary, the temporal effect of air turbulence results in continuous wavefront aberration in the pupil plane and corresponding distortion or shifting of the PSF. The resulting image is thus formed as a distorted or shifted PSF and the original object, with additive noise due to data acquisition. These aspects are related to the temporal imaging model and are discussed in the following section.

3.7.1 The spatiotemporal PSF

In developing an imaging model we need to recognise that the quantities involved are time variant: the object and point spread function and therefore the image also, are all functions of time. For completeness therefore, we should write

$$g_1(p,q,t) = \sum_{k=1}^{N}\sum_{l=1}^{M} f(k,l;t)h(p-k,q-l;t) + \eta(p,q;t), \tag{3.42}$$

for the time-variant, spatially-invariant image model, and,

$$g_2(p,q,t) = \sum_{k=1}^{N}\sum_{l=1}^{M} f(k,l;t)h(p,q;k,l;t) + \eta(p,q;t), \tag{3.43}$$

for the time-variant, spatially-variant image model.

The image model used to represent the work outlined in this thesis is shown in Figure 3.8. The input-output relationship of an object, $f(k,l;t)$, can be expressed as a convolution

with a degradation function $h(p,q;k,l,t)$, with additive noise source, $\eta(p,q;t)$. This result is measured in the spatio-temporal domain, i.e., $g_2(p,q;t)$, as defined by Equation 3.42. Models of this kind are known as the *forward problem* [8].

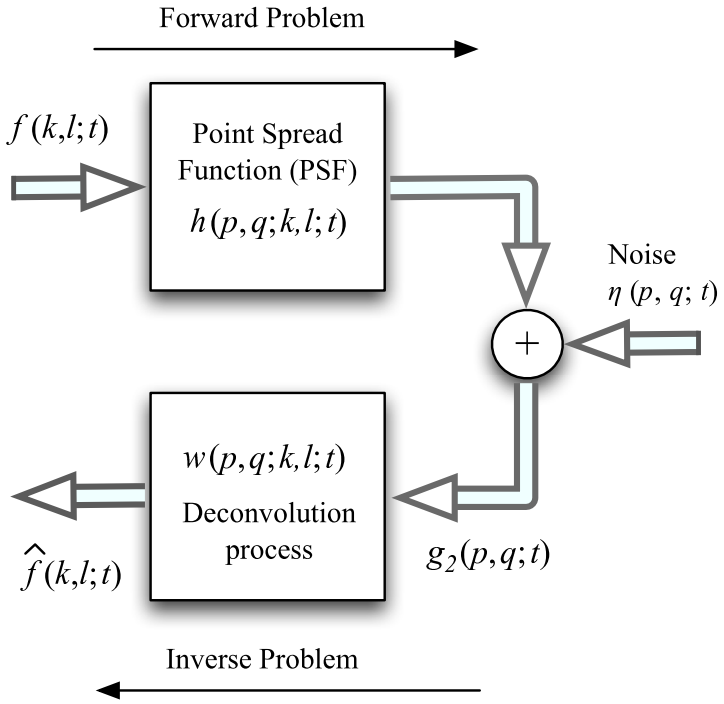

Figure 3.8 Generalised imaging model.

The recovery of the original image, $f(k,l;t)$, in the presence of noise, $\eta(p,q;t)$, typically results in an estimation of the original image, $\hat{f}(k,l;t)$. Restoration problems such as this are referred to as the *inverse problem* and are typically ill-posed. To restore $f(k,l;t)$, knowledge of the PSF, i.e., $w(p,q;k,l;t)$, is usually required. Such knowledge can be *a priori*, as in the case of blind deconvolution [7], or *a posteriori* as in the case for deconvolution from wavefront sensing (DWFS). In this study, the latter is used, where an estimate of the PSF is obtained from wavefront sensing data. Various methods of deconvolution, employing the spatiotemporal PSF defined here, are explored in Appendix A.

3.7.2 Spatial variance of the anisoplanatic PSF

As shown earlier in this section, the spatially-invariant PSF (SIPSF) degradation function is independent of position in the image plane. An example of this is the occurrence of a sudden camera movement during an exposure. The resulting blur degradation (defined in terms of a single PSF) is applied over the entire image plane. The resulting image, $g(p,q)$,

is a convolution of the object, $f(k,l)$, with the blur function, $h(p,q;k,l)$, and in the absence of noise this is given by Equation 3.38.

However, wide-field astronomical images are of particular interest in this study. Since the pupil plane image is anisoplanatic over a wide FOV, the PSF will be spatially variant (SVPSF). In such cases the PSF is independent over the image plane, i.e., it varies spatially. An example of the SVPSF is taking an exposure out of a moving train, where the lower part of the image might be composed of the internal carriage (possibly out of focus), and the upper portion of the image would be a view outside of the carriage. If the train is moving rapidly past a nearby object, such as a signal station, after a long exposure the upper portion of the resulting image would show a motion blurred object outside the carriage. However, the lower portion of the image (the carriage) would be independent of motion blur. Thus, for this example, two degradation functions would be required to restore the entire image, one that incorporates a motion de-blurring function applicable to the top portion of the image, another, possibly defocus, for the lower portion of the image.

For this study, the discrete spatiotemporal model described by Equation 3.43, was used to represent an anisoplanatically blurred image over time. Recovery of the SVPSF is based on modal expansions (Zernike coefficients) described in Section 3.3. The method used to recover the Zernike coefficients over a wide field for reconstruction of the anisoplanatic, spatiotemporal PSF, is detailed in Chapter 6.

Chapter 4

Adaptive Optics

Adaptive optics is a technology that utilises optical and control engineering to remove temporally and spatially varying optical wavefront abberations, caused by the propagation of light through a distorting medium such as the Earth's atmosphere [134].

Atmospheric turbulence severely limits the performance of ground-based telescopes; to astronomers, this results in a deterioration of *seeing* conditions. Adaptive optics (AO) is used on ground-based telescopes to ultimately provide diffraction limited images. Currently, the achievement of diffraction limited resolution by ground-based AO telescopes is limited to infrared wavelengths [27]. The Hubble Space Telescope (HST), on the other hand, is capable of resolving images close to the diffraction limit over a wide electromagnetic spectrum, i.e., from ultraviolet to near infrared wavelengths.

In spite of this current limitation, the refinement of AO systems has continued to improve [90]. In some areas of research, the infrared performance of ground based telescopes using *extreme* adaptive optics, exceeds the performance of the Hubble Space Telescope (HST) [27]. Advances, in terms of sensor technology [107], improved control algorithms [24], and, more recently, wavefront prediction [90], are demonstrations of further refinements to AO systems. New fields in adaptive optics, such as Multi-Conjugate Adaptive Optics (MCAO) [74] and Wide Field Adaptive Optics (WFAO) [32], will play a significant part in extending the application of AO over the next decade.

An assessment of the efficiency of current AO systems was of interest, and served as a basis for qualitative measurement used in related research. Methods to assess AO system performance and a comparative study have been developed by Roddier [115]. A more re-

Table 4.1 Adaptive Optics performance comparison [27, 120].

Telescope or AO system	**Altitude** (m)	**Seeing** (arcsec.)	**AO resol.**[a] (arcsec.)	**Diffraction limit (DL)** [a] (arcsec.)	**Strehl ratio** [a] (DL$\geq$ 0.8)
HST	593000	—	—	0.050	> 0.80[b]
PUEO AOB	4200	0.55	0.20	0.126	0.50
ChAOS	3000	1.00	0.40	0.050	0.50
ADONIS	2400	0.80	0.34	0.226	0.40
ACE - ADOPT	1750	1.20	0.13	0.070	0.40
Lick AO	1300	1.50	0.13	0.130	0.65
Yerkes	235	2.00	0.80	0.533	0.30

[a] K-Band
[b] SR = 0.98 at 1.2 μm

cent study compares the performance of several ground-based AO system and includes the Hubble Space Telescope (HST) as a benchmark [120]. A summary of these results, assuming on-axis operation, is presented in Table 4.1. The term *on-axis* is defined here as either direct measurement of phase perturbations from a science object, or indirect measurement using a guide star within the isoplanatic *patch* [134], defined by θ_0.

The second column in Table 4.1 is ranked by altitude as this is a significant factor in assessing the performance of an AO system. For example, the Hubble Space Telescope (HST) orbiting Earth at an altitude of 593 km is an imaging instrument capable of optimal performance over a wide electromagnetic spectrum (between infrared and ultraviolet wavelengths). The atmosphere, however, acts like an opaque screen to ultraviolet radiation; attenuation in the visual spectrum is largely dependent on wavelength. Correcting for wavefront aberrations in the visual spectrum is a significant challenge for AO researchers.

The third column in Table 4.1 compares the *seeing conditions* of several ground-based telescopes operating at various altitudes; due to the high altitude of operation, this parameter is not applicable to the HST. Improvement, in terms of angular resolution for ground-based systems, is shown in the 4^{th} column with the application of an AO system. A comparison of columns 4 and 5 shows the effectiveness of AO for K-Band frequencies on ground-based telescopes, where the term *diffraction limit* is defined by the Rayleigh criterion [135] and states that the angular resolution of an instrument such as a telescope with a circular

aperture D and wavelength λ, is limited by $1.22\lambda/D$ radians.

The last column on the right of Table 4.1 provides a relative performance measure for each system. A useful performance measure is the Strehl ratio (SR) [49]. Strehl is a measure of the performance of the AO system, relative to the diffraction limited performance of the telescope optics. A ratio between a diffraction limited PSF and an aberrated PSF is the basis for this measure; a formal definition of the Strehl ratio is given in Section 4.3.

A comparison of several AO systems, taken from data presented in Table 4.1 but limited to ground-based telescopes, is shown in Figure 4.1. In addition to attenuation of visual

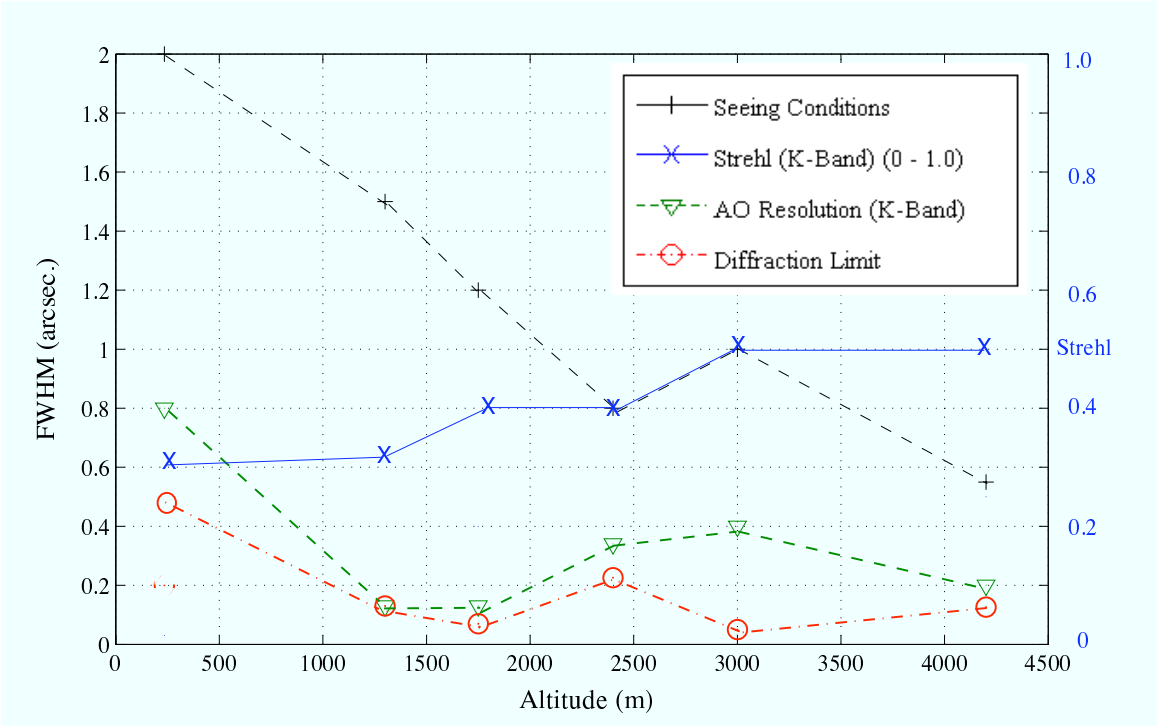

Figure 4.1 Performance of the AO systems listed in Table 4.1.

and ultraviolet radiation due to the atmosphere, ground-based astronomical instruments are severely effected by the sensitivity of the foreground radiation of the Earth's atmosphere. For example, the Hubble space telescope has eight times more contrast than the sky on a moonless night at Manua Kea (4200 meters) [27]. However for the purposes of this comparative study, performance for K-band wavelengths (within the 2.2 μm range), are used.

Figure 4.1 plots the diffraction limit (DL) of six AO systems listed in Table 4.1; since this

theoretical limit is based on the optics of the instruments, and is independent of environmental or turbulence conditions of the site, this figure is relatively consistent with most medium to large telescopes of apertures, $1 \leq D \leq 10$ m. For example, the DL for the McLellan telescope at MJUO, given $\lambda = 2.2\ \mu$m and $D = 1$ m, is 0.55 arcseconds. However, seeing is inversely proportional to altitude. When an AO system is installed and is fully operational, the AO resolution can be seen as an effective measure to counter adverse seeing conditions and approach the optical resolution limits of the instrument. Depending on the AO system, and increased constraints with respect to the operational environment, diffraction limited performance can be achieved at micrometer wavelengths.

In summary, the performance of optical telescopes fitted with AO systems is constrained by:

1. Air density: atmospheric turbulence resulting in poor seeing conditions is proportional to altitude; the density of air however, is inversely proportional to altitude.
2. Wavelength: longer, infrared wavelengths are less attenuated by the atmosphere than shorter wavelengths, such as in the visual spectrum. AO systems can attain diffraction limited performance, however such performance is currently limited to K-Band (2.2μm) wavelengths.
3. On-axis measurements: bright guide stars are limited to $< 1\%$ of the sky [27]. Performance of AO systems degrade as off-axis measurements of phase perturbations increase.
4. Foreground radiation: the Earth's foreground radiation is reduced with altitude, resulting in more contrast at higher altitudes. This condition can have a significant affect on increasing the signal-to-noise ratio (SNR) in astronomical imaging.

Considering the above constraints, Item-3 has particular significance to this study.

The reminder of this chapter is organised as follows. A brief background on atmospheric turbulence is given in the next section. Parameters, based on the *atmospheric imaging problem*, are also established; these are used to develop a propagation model described in Chapter 6. The temporal behaviour of Zernike modes are analysed in Section 4.2 to determine predictability. Specifically, low-order Zernike coefficients, representing phase perturbations from imaging stellar objects through turbulence, are defined in terms of wavefront maps. A series of temporally evolving wavefront maps result in time series data used in

this study. In Section 4.3 the effects of imaging through turbulence are reviewed. Results from simulations comparing Kolmogorov spectra to a spatiotemporal wavefront model are presented. A new method to extend the curvature and geometric wavefront sensors for use with multiple source objects is described in Section 4.4, and various deconvolution from wavefront sensing models are reviewed in Section 4.5. These methods are used for acquisition of empirical data throughout this dissertation, and in simulations detailed in Chapter 6. Section 4.6 reviews current methods to estimate the anisoplanatic PSF. Lastly, Section 4.7 discusses wavefront deformation in the metapupil, in terms of several turbulence models and qualification of Taylor's hypothesis over large field angles.

4.1 Atmospheric turbulence

The *thin layer* atmospheric model was briefly described in the introductory chapter. In this section, mathematical preliminaries used to describe atmospheric turbulence are summarised and the effects on the anisoplanatic PSF are given. Parameters defined in this chapter are used to characterise a simulation environment described in Chapter 6.

4.1.1 Characterising air turbulence

The refractive index of air is close to unity but small variances have a significant effect on imaging. Refractivity at optical wavelengths is proportional to P/T, where P is pressure in millibars and T is temperature in degrees Kelvin. Since the optical effects of imaging through turbulence are a consequence of incomplete mixing of air at different temperatures, air turbulence can be characterised in terms of altitude and time [49]. This relationship is given in the form of the refractive index structure constant.

The refractive index structure constant

The refractive index structure constant, C_n^2, is used to characterise the strength of atmospheric turbulence and can be expressed in terms of the height of a turbulent layer, h, and the velocity, v, at which it travels. C_n^2 profiles are used as a basic parameter in subsequent mathematical expressions.

As a consequence of individual site characteristics, several C_n^2 models have been developed.

A popular model is the Hufnagel-Valley turbulence profile [117], given by

$$\begin{aligned} C_n^2(h) =& 5.94\times10^{-53}(v/27)^2h^{10}\exp[-h/1000]+ \\ & 2.7\times10^{-16}\exp[-h/1500]+A\exp[-h/100], \end{aligned} \tag{4.1}$$

where A provides a characteristic for near ground turbulence, and v is the high altitude wind velocity. Commonly used values for A and v are 1.7×10^{-14} $\mathrm{m}^{-2/3}$ and 21 $\mathrm{m\,s}^{-1}$, respectively. A $C_n^2(h)$ model and velocity profile, $v(h)$, have been developed for Mt. John University Observatory (MJUO) in New Zealand [91]. Relating the strength of atmospheric turbulence to an effective aperture diameter was achieved by Fried [30] and is referred to as the Fried coherence length parameter, r_0.

The Fried coherence length

Many of the parameters used to characterise atmospheric turbulence are formulated from the Fried coherence length parameter, r_0. The Fried parameter is the effective aperture size, given the strength of turbulence in terms of the refractive index structure constant C_n^2, beyond which an increase in the aperture diameter does not increase resolution. Fried's parameter is given by [117],

$$r_0 = 0.185\left[\frac{4\pi^2}{k^2\int_0^L C_n^2(z)}\mathrm{d}z\right]^{3/5} \tag{4.2}$$

where $k=2\pi/\lambda$, L is the propagation distance through turbulence, and C_n^2 is the structure constant; a profile of which is given by Equation D.16.

4.1.2 The Kolmogorov turbulence model

As described in Chapter 1, the atmosphere supports two or three thin turbulent layers, where each layer can be likened to a collection of lenses of varying refractive indexes, moving at different velocities with respect to a fixed, ground-based aperture. Until relatively recently however, such a depiction of the atmosphere lacked mathematical foundation.

For over forty years the Kolmogorov statistical model [72] has been used to define the spatial structure of turbulent air flows [30]. Extending Kolmogorov's work to include the statistics of refractive index fluctuations was made by Tatarski [127]. The model has the unusual property that it appears the same, irrespective of the scale used. This implies that the model is based on a fractal property, characteristic in Nature [122].

To describe the Kolmogorov turbulence model we need to define both the temporal and spatial structure of the atmosphere in terms of random processes. In this section, the spatial statistical model is discussed; temporal considerations, such as determining the period for decorrelation of the atmosphere, are addressed in the next section.

Firstly, consider the refractive index variations of two points in 3D space, $\mathbf{r}$ and $\mathbf{r}+\mathbf{r}'$. The covariance function given by Equation 2.18 can be used to describe the refractive index variations for a stationary process as

$$C_n(\mathbf{r}') \equiv \big\langle (n(\mathbf{r}-\langle n\rangle)\,(n(\mathbf{r}+\mathbf{r}')-\langle n\rangle)\big\rangle. \tag{4.3}$$

where $n(\mathbf{r}) = \mathscr{X}$ and $n(\mathbf{r}+\mathbf{r}') = \mathscr{Y}$ from Equation 2.18.

Considering the covariance function given by Equation 4.3, the structure function is simply the mean square difference between two random processes,

$$\begin{aligned} D_n(\mathbf{r}') &= \big\langle |n(\mathbf{r}) - n(\mathbf{r}+\mathbf{r}')|^2 \big\rangle \\ &= 2\big[C_n(0) - C_n(\mathbf{r}')\big]. \end{aligned} \tag{4.4}$$

Now consider a phase shift at point $\mathbf{r}$; the function $\phi(\mathbf{r})$ is the result of random refractive index fluctuations given by

$$\phi(\mathbf{r}) = k\int_h^{h+\delta h} n(\mathbf{r},z)dz, \tag{4.5}$$

where k is the wave number, h is the height of the layer, δh is the layer thickness, and z is the altitude at zenith.

Such phase shift fluctuations can be described in terms of a structure function, using the same formulation as given by Equation 4.4 for refractive index variations. Thus, the structure function for phase shift fluctuations can be expressed as [117]

$$D_\phi(\mathbf{r}') = \big\langle \big[\phi(\mathbf{r}) - \phi(\mathbf{r}+\mathbf{r}')\big]^2 \big\rangle, \tag{4.6}$$

where $D_\phi(\mathbf{r}')$ represents the spatial structure function for phase variations and $\phi(\cdot)$ is a function of the refractive index fluctuations over a spatial separation of $\mathbf{r}$ and $\mathbf{r}+\mathbf{r}'$.

Fried showed that the phase structure function could be equated in terms of coherence

length, r_0, [29]

$$D(\mathbf{r}) = 6.88\left(\frac{\mathbf{r}}{r_0}\right)^{5/3}, \tag{4.7}$$

where the Fried coherence length, r_0, is defined by Equation 4.2.

The power spectrum for Kolmogorov turbulence, $\Phi_n^K(\boldsymbol{\kappa})$, provides a statistical distribution of the size and number of turbulent eddies and uses an independent variable, $\boldsymbol{\kappa}$, as a spatial wavenumber vector, where $\boldsymbol{\kappa}$ is defined within the range, $2\pi/L_0 \leq \boldsymbol{\kappa} \leq 2\pi/\ell_0$. Since the power spectrum $\Phi(\boldsymbol{\kappa})$ of a stationary process is the Fourier transform of the covariance function,

$$\Phi(\boldsymbol{\kappa}) = \int_{\infty}^{\infty} C(\mathbf{r}') \exp[-2\pi i \boldsymbol{\kappa}\mathbf{r}'] d\mathbf{r}', \tag{4.8}$$

the power spectrum of the refractive index fluctuations is given by

$$\Phi_n^K(\boldsymbol{\kappa}) = 0.033\, C_n^2(z) |\boldsymbol{\kappa}|^{-11/3}. \tag{4.9}$$

A fundamental characteristic of Kolmogorov turbulence is that the inner and outer scales of turbulence, $\{L_0, \ell_0\}$ are bound; this can range from a few millimeters to several meters. Thus, the statistical nature of a portion of the atmosphere can be defined using these basic formulations [117].

Extending the outer scale of this statistical model results in a progressively more complex structure and unpredictable, random motion of air. This is generally referred to as non-Kolmogorov turbulence.

4.1.3 Temporal decorrelation

Defining the minimal temporal frequency is a requirement to ensure that the effects of turbulence, as it moves over the aperture propelled by wind, is suspended. The general assumption is that low-order perturbations, such as tilt, move more slowly that high-order eddies; this is primarily due to the principle of inertia [134]. Thus, in terms of correcting either low- or high-order aberrations, two expressions are used to determine the minimum sampling frequency to ensure temporal correlation of the atmosphere during short exposures.

For tilt only compensation, f_T is given by Tyler [133], however this is known as the *tilt*

Greenwood frequency [134]

$$f_T = 0.33D^{-1/6}\lambda^{-1}\sec^{-1/2}\beta\left[\int_{Path} C_n^2(h)\,v_w(h)dh\right]^2, \tag{4.10}$$

where λ is the wavelength, h is the altitude of the turbulent layer, v_w is the wind velocity, C_n^2 is the turbulence structure constant,and β is the angle of observation from zenith.

To effectively *freeze* the effects of higher order aberrations, the Greenwood frequency, f_G, is used and is given by [45],

$$f_G = 2.31\lambda^{-6/5}\left[\sec\beta\int_{Path} C_n^2(h)\,v_w(h)^{5/3}dh\right]^{3/5}. \tag{4.11}$$

For this study, the acquisition of both low and higher order aberrations, in terms of Zernike coefficients $a_2 \cdots a_{11}$, were required. However, since the placement of source beacons and target objects for simulations were near zenith, and a constant wind velocity was assumed, a simplified version of Equation 4.11 was used [134]

$$f_G = 0.43\left(\frac{V_{wind}}{r_0}\right), \tag{4.12}$$

where r_0 is Fried's coherence length given by Equation 4.2, and V_{wind} is a constant wind velocity near zenith.

4.2 Time series analysis of Zernike modes

Temporal studies have been conducted on low-order Zernike coefficients representing tilt aberrations from imaging stellar objects through turbulence [34, 83, 88]. Insight into the temporal behaviour of atmospherically distorted wavefronts has been used in simulations to provide temporal spectra in accordance with theoretical predictions [61]. As a consequence of these studies, questions have been raised concerning the predictability of wavefront perturbations.

Consider, for example, a set of signals comprising the output of a wavefront sensor and represented by a set of Zernike coefficients. If such a set is deterministic, it may also be predictable [90]. The objective here is to firstly establish the predictability of a range of Zernike modes, and secondly, use these modes to train a system to respond to the predictable portion of corresponding spatiotemporal patterns, thus allowing the temporal decorrelation

of the wavefront to be partially negated. An ESN is proposed to achieve this, resulting in prediction of the SVPSF. In contrast to state space modelling techniques, such as the extended Kalman filter, the ESN supports a simplified training methodology that can be adapted to learn the short-term evolution of a set of Zernike coefficients.

4.2.1 Background

A signal analysis was conducted on the temporal behaviour of low-order Zernike coefficients, $\{a_2, a_3\}$, by McGaughey and Aitken [88]. For their study, rescaled-range (RS) analysis and a correlation algorithm were used to detect the dependence on samples within a time series and the presence of deterministic chaos, respectively. Their work, using a H-S WFS, concluded that predictability of wavefront tilt resulted from the spatial averaging performed by the WFS [90].

Wavefront prediction has helped reduce temporal decorrelation in adaptive optics (AO) systems. For example, due to finite system response times, lag compensation is often required in AO control systems [94, 105]. Extended Kalman filters have traditionally been applied, however wavefront tilt has been predicted 30 time-samples into the future using an artificial neural network [73]. While such networks are effective in closed loop systems, the literature is sparse on direct use of such predictions for open loop system applications, e.g., deconvolution from wavefront sensing.

By assessing the properties of a time series using statistical methods and temporal analysis [17], the underlying dynamics of a system can be classified as either deterministic and possibly chaotic, or non-deterministic. In practice, however, most systems are a combination of behaviours [123]; determining the proportion of each, especially considering the effects of additive noise, is seldom definitive. Signal processing methods, such as detrending, filtering, etc., are often applied to time series data to partially recover low SNR sequences. However, specialised processes are often required to unravel the layers of embedding dimensions, characteristic of chaotic signals, and to reveal internal structure. Representing complex dynamics with simple models is a motivation to assess the presence of chaos in a system [123].

Atmospheric turbulence has a deterministic component that can be predicted [90]. Using the centroids of a star's position from an H-S WFS, it has been shown that the chaotic temporal trajectory has a correlation dimension, $D_r \approx 6$ [65]. The analysis conducted in this subsection explores the possibility of predicting higher order Zernike modes.

4.2.2 Acquisition

A raw image dataset of a single NGS, `NGS6a15_5`, was acquired on 15 June 2007 at MJUO using a 1-m class catadioptric telescope. The dataset comprising 5000 frames, was recorded with good seeing at a sampling frequency, $f_s = 100$ Hz. A series of corresponding wavefront maps were retrieved and each map generated 19 Zernike ensembles, $\{a_2, a_3, \cdots, a_{20}\}$, at the output of an extended geometric WFS. Acquisition details for this, and several other related datasets, are listed in Table C.4 of Appendix C.

Of these ensembles, wavefront tilt, defocus, and astigmatism, representing Zernike modes $\{a_2, a_3\}$, a_4, and $\{a_5, a_6\}$, respectively, were selected for time series analysis. Examples of temporal ensembles for tilt and astigmatism are shown in Figure 4.2 (a) and (b), respectively. Tilt components were selected as they could be compared with the results from a previous study by McGaughey and Aitken [88]. The remaining random signals, i.e., defocus, and astigmatism components, were selected to ascertain the extent of individual predictability relative to lower-order modes.

A visual inspection of the responses shown in Figure 4.2 did not reveal any underlying properties, or correlation between data points or sequences, e.g., no periodic characteristics were noted. However, the amplitudes of the modes were steadily reduced as a function of mode due to the shape of the $f^{-8/3}$ Kolmogorov spectrum [48].

To evaluate the characteristics of deterministic chaos, time series analysis was conducted on five Zernike modes using the public domain time-series analysis tool, TSTool, [104].

4.2.3 Analysis

Five time series ensembles, representing Zernike coefficients $\{a_2, a_3, \cdots, a_6\}$, were generated from field data `NGS6a15_5`. Each ensemble was analysed to determine the extent of low-dimensional chaotic attractors and to assess prediction error. Nonlinear time series analysis was employed, facilitated by the package, TSTool. For reasons that will be outlined, three time series parameters were of particular interest. These parameters are:

- The minimum embedding dimension[1], D_E.
- The largest Lyapunov exponent, λ_1.

[1] The minimum embedding dimension is determined using Cao's method [15].

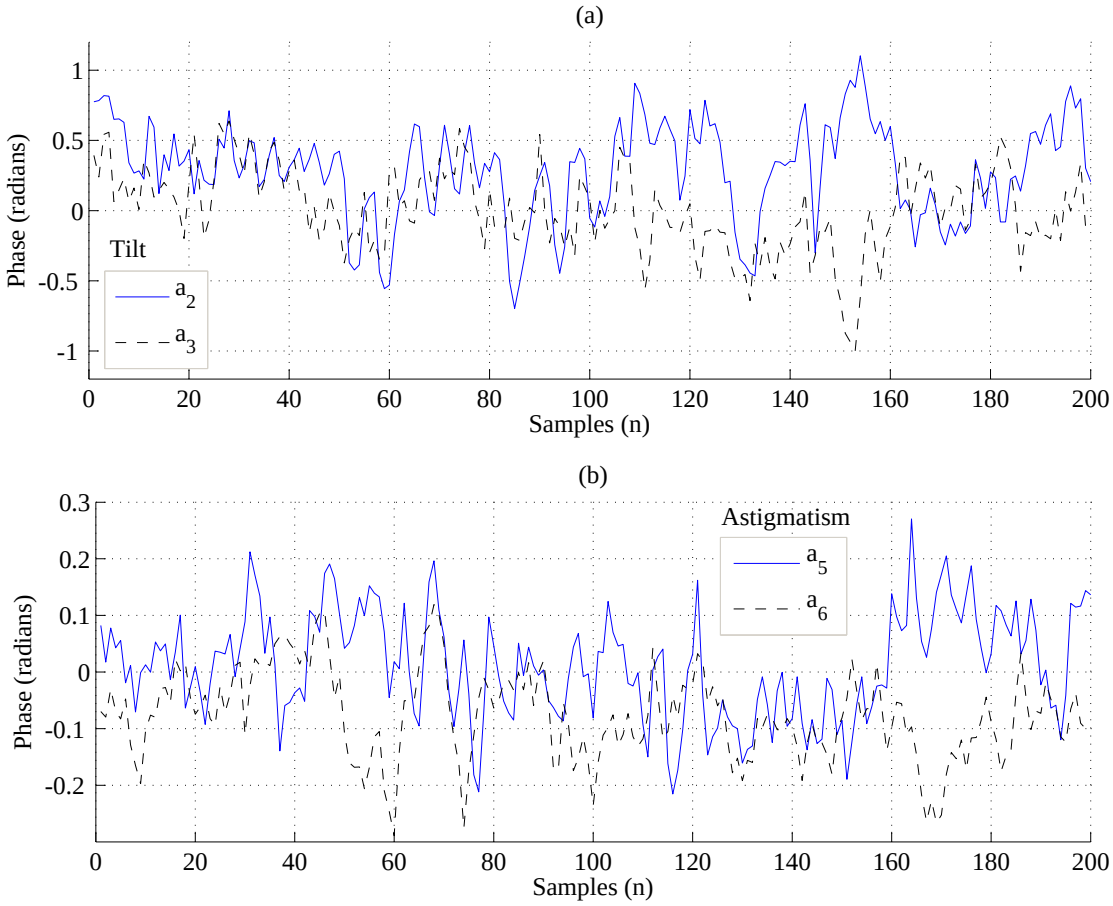

Figure 4.2 Time-series plots of Zernike coefficients from dataset, NGS6a15_5 (a) Tilt terms (a_2, a_3), (b) Astigmatism (a_5, a_6). The sampling rate is 100 per second.

- Correlation dimension, D_C.

The minimum embedding dimension, D_E, is the minimum dimension required to provide an underlying structure of data. When $\mathrm{D}_E < 2$, the trajectory of a series can be determined using state space methods. However, non-integer dimensions, where $\mathrm{D}_E > 2$ suggest that the system is chaotic; when $\mathrm{D}_E = 2$ the trajectory would lie on a surface and could not interact with itself [123].

Chaotic behaviour can be determined by analysing the Lyapunov exponents, λ_k, where $k \in \{1, 2, \cdots, K\}$, since these represent the dynamics of a time series. Lyapunov exponents are closely related to eigenvalues; however, while the latter are typically used to represent equilibrium points in state space, the former are geometrically averaged along the orbit or trajectory of a time series signal [123]. A dynamical system with K dimensions has K Lyapunov exponents, and K eigenvalues at each point [123].

Table 4.2 Time series analysis - parameter summary.

Time series parameter	Zernike coefficient				
	a_2	a_3	a_4	a_5	a_6
Embedding dimension, D_E	4.1	4.2	5.4	6.1	6.2
Lyapunov exponent, λ_1	0.32	0.39	0.41	0.417	0.423
Correlation dimension, D_C	3.0	3.1	3.62	3.9	4.0

To estimate the Lyapunov exponents, TSTool uses an algorithm similar to Wolf [146]. The average exponential growth of the distance of neigbouring orbits is calculated from the time series by firstly estimating prediction error. An estimate of the largest Lyapunov exponent is based on increases in the prediction error, over the number of iterations. In the following analysis, only the largest Lyapunov exponent, λ_1, is evaluated.

The correlation dimension, D_C, is evaluated, since it provides a measure of chaotic behaviour within a system. This parameter represents the slope of the correlation sum, $C(r)$, defined as the number of data points that revisit a series of radii, r, over the limit of N data points in a series, and the limit of small r. This is given as, [123]

$$D_C = \lim_{r\to 0}\lim_{N\to\infty} \frac{d\log C(r)}{d\log r}, \tag{4.13}$$

where $C(r)$ is defined as,

$$C(r) = \frac{2}{N(N-1)} \sum_{j=1}^{N} \sum_{i=j+1}^{N} \Theta(r - r_{ij}), \tag{4.14}$$

and where $\Theta(\cdot)$ is the Heaviside function, defined for x as,

$$\Theta(x) = \begin{cases} 1 & \text{for } x < 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{4.15}$$

The correlation dimension tests the presence of chaos and has been reported in the literature for wavefront tilt [88]. The results of time series analysis conducted using TSTool on empirical data are presented in Table 4.2.

Table 4.2 provides a basis for further discussion. Firstly, a general observation of these results showed that all Zernike ensembles were consistent in terms of the three parameters used to test predictability; based on these data, if predictability can be qualified for one coefficient ensemble, then the remaining series must also qualify.

Secondly, evidence of a low-dimensional attractor is a strong indication of a chaotic process [123]. Independent, identically distributed and coloured noise processes are stochastic and have an infinite number of degrees of freedom, however processes that can be identified as chaotic usually have an embedded dimension, $D_E < 10$ [88]. This result was recorded for each Zernike ensemble tested in this analysis.

Thirdly, the sign of the largest Lyapunov exponent, λ_1, qualifies chaos, and the value quantifies the extent of chaotic behaviour [123]. For example, given a bounded dynamical system that supports a positive Lyapunov exponent, the system can be classified as chaotic, and the value of the exponent describes the average rate at which predictability is lost [123]. The results in Table 4.2 show a strong indication of chaotic behaviour. Further, the increasing value of the exponent for Zernike modes a_2 to a_6, suggests that higher modes are more chaotic, and therefore more difficult to predict than lower modes.

Lastly, the correlation dimension, D_C, is a good indication of chaotic behaviour, as it emphasises the regions of the attractor visited most frequently by the orbit [123]. The values shown in Table 4.2 are consistent with other parameters, i.e., they show increasing chaos as a function of Zernike mode. The correlation dimension for low-order tilt was compared to a similar analysis conducted by Jorgenson and Aitken, where $D_C \approx 6$ [65].

In summary, the time series analysis outlined in this section suggests that each modal aberration tested has a deterministic component that could be predicted. Several studies on the prediction of wavefront tilt using artificial neural networks have been conducted [22,76,94]. A conclusion reached by McGaughey and Aiken [88] was that empirical wavefront data is well described as filtered, fractal Brownian motion, and does not show signs of chaotic behaviour. However, their study also showed that low-pass spatial filtering appeared to be responsible for the persistence and predicability found in time sequences of wavefront slopes.

This analysis has shown that firstly, predictability of wavefront tilt, in addition to higher order aberrations, is present in data conversions measured using a geometric wavefront sensor. The suggestion that the geometric wavefront sensor contributed to these chaotic properties

was not determined in this study, however this comprises future research. Secondly, the temporal analysis of higher order Zernike terms suggested that chaotic behaviour increased with modal order.

4.3 Imaging through turbulence

As shown in the introduction of this dissertation, atmospheric turbulence severely affects the resolution of ground-based astronomical telescopes and distorts images of exoatmospheric objects. Propagation of electromagetic waves through the atmosphere, originally in the form of planar wavefronts, undergo a phase distortion. Such phase distortion also introduces amplitude fluctuations, referred to as scintillation. Since light from astronomical objects is incoherent, aperture averaging occurs when such objects are viewed through a telescope [134]. As a result, the intensity variations are small in comparison with phase perturbations.

The adverse effects of turbulence on point-source astronomical objects is to spread light over the image plane. For example, the PSF or Airy pattern, characteristic of a diffraction limited image [12], is either distorted and/or displaced in the exit pupil. Effects of low-order aberrations such as tilt (Zernike polynomial orders Z_2 and Z_3) displace the PSF, however higher orders ($Z_4 \cdots Z_\infty$) aberrate the PSF; some high order aberrations such as the x- and y-component of Coma (Z_7 and Z_8) also contribute to displacement of the PSF. Such wavefront aberrations are shown in Figure 3.4.

The effects of wavefront phase aberrations, in terms of quantifying the effects of wavefront error on imaging, and the subsequent definition of relevant parameters, are summarised in this section and form the basis of discussion throughout this thesis.

4.3.1 Anisoplanatism

A source beacon, e.g., a guide star, and a target or science object, such as a faint ($m_v > 16$) galaxy, are not affected by the same patch of atmospheric turbulence. The angular separation between the object and the guide star is bounded by the isoplanatic angle, θ_0 [101]. The effects of turbulence over a region, referred to in the literature as the *isoplanatic region*, and as defined by the isoplanatic angle, may be considered to be constant [30]. The size of the isoplanatic region over the image plane is defined by θ_0, as it subtends a portion of the FOV from the centre of the guide star. An anisoplanatic region is that which is not

inclusive of the isoplanatic region.

The isoplanatic angle as given by Fried [30] is

$$\theta_0 = 58.1 \times 10^{-3} \lambda^{6/5} \Big[\int_0^L C_n^2(z) z^{5/3} dz \Big]^{-3/5} \text{ rad}, \tag{4.16}$$

where λ is the optical wavelength, z is the altitude, C_n^2 is the structure constant of the turbulence, and L is the path length through turbulence.

To quantify the error induced by a PSF as it explores image space, a simulation was conducted using a phase screen [47]. A point-source was projected through the phase screen and moved over an anisoplanatic region. The simulation generated a distortion matrix that represented the spatially variant PSF (SVPSF). An $N \times N$ isoplanatic sub-region, comprising a reference PSF, $h_{Ref}(\cdot)$, and centred within a $X \times Y$ anisoplanatic region was used as the basis for comparison. As the angular separation and orientation of each PSF to the reference PSF was varied, a 2D array of SIPSFs, $h(\cdot)$, was calculated. Each calculation used $N \times N$ sized regions and was based on phase distortions at each x- and y-coordinate, $\phi(x,y)$. As the SIPSF, $h(\cdot)$, explored both isoplanatic and anisoplanatic regions, the mean square error (MSE), based on the reference SIPSF, $h_{Ref}(\cdot)$, was calculated. These data resulted in a 2D error function over the anisoplanatic region, $\varepsilon^2(x,y)$.

The expression used to estimate the MSE of the SVPSF using a single source object for reference, is given by

$$\begin{aligned} \varepsilon^2(x,y) = \frac{1}{N^2} \sum_{k,l \in \Gamma} \Big[& h(k+x, l+y) - \\ & h_{Ref}(k+X/2, l+Y/2) \Big]^2. \end{aligned} \tag{4.17}$$

A 2D simulation of the SVPSF is shown in Figure 4.3. The resulting PSF is spatially variant, since the perturbing medium is inhomogeneous, i.e., the 2D plot of the resulting SVPSF would differ if the source point $p(x,y)$ were moved w.r.t. the perturbation. A generalised version of Equation 4.17, but where multiple source and target objects are employed, is described in Appendix A.3.

The significance of this simulation demonstrates that for severe degradations, even the smallest deviation from a point source reference beacon can result in significant wavefront errors. Such errors directly result in image distortion, preceding the application of restora-

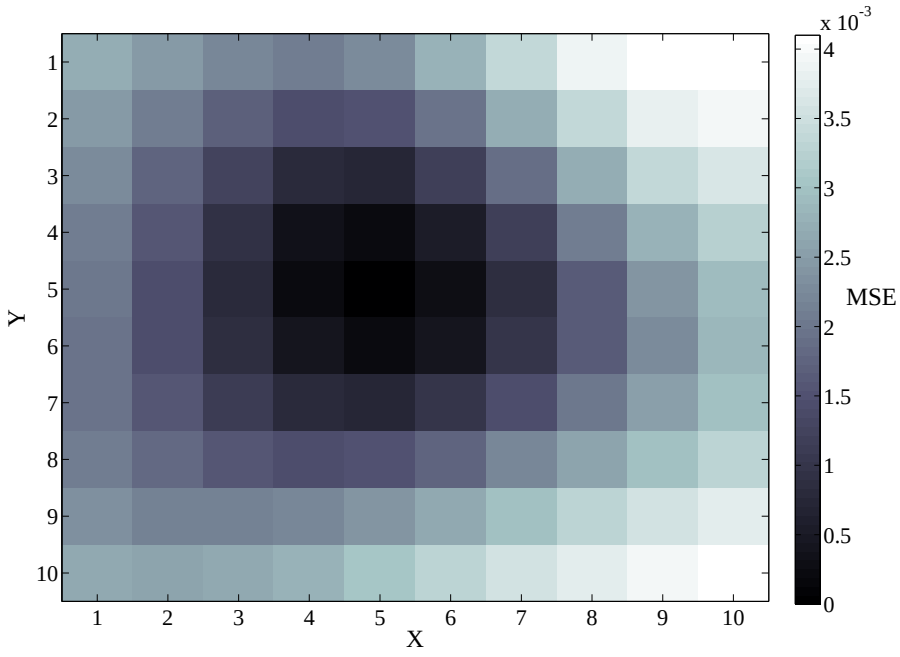

Figure 4.3 MSE error plot of the spatially variant PSF.

tive algorithms. Given this simplified simulation model, the isoplanatic angle, θ_0, was estimated to be 10 μrads. This result is in agreement with Roggemann and Welsh [117].

4.3.2 The Taylor frozen turbulence hypothesis

The Taylor frozen turbulence hypothesis was proposed by G. I. Taylor in 1938 and was a practical attempt to simplify the analysis of turbulence passing over an aperture [128]. Specifically, the hypothesis states that turbulent air moving over the imaging path of an optical instrument, with winds moving the turbulence at velocity v_s, has insufficient time to change, as viewed through the aperture, i.e., the turbulence is simply displaced. Essentially, the turbulence remains *frozen* for a period, τ_T, which is less than the time required for the turbulence to pass over the aperture.

This suggests that in terms of a short exposure image, the distortion field is maintained and simply displaced over the aperture for a period of τ_T. As *frozen* turbulence moves over the aperture, it is replaced by a new perturbation field.

The Taylor hypothesis can be expressed as $v_s(t_2 - t_1)$ and is related to the space-time cor-

relation function of the atmosphere by [117]

$$\Gamma_{p_i}(\Delta\mathbf{x}, t_2, t_1) = \langle \exp[j\psi_i(\mathbf{x}, t_1) \exp[-j\psi_i(\mathbf{x} - \Delta\mathbf{x}, t_2)] \rangle, \tag{4.18}$$

where $\Delta\mathbf{x}$ is the spatial separation between two points in the pupil, and ψ_i is the phase perturbation in the i'th layer.

Goodman listed two criteria to ensure quality, multiple frame image captures [43]

1. Consistent atmospheric conditions during the period of integration.
2. Prior knowledge of the spatial frequencies.

Given isoplanatic imaging applications, these conditions can be combined and expressed in terms of the temporal decorrelation of the atmosphere. The Greenwood frequency can be used to express the bandwidth specification for an adaptive optics system [45]; i.e., if the calculated frequency is exceeded, turbulence will not evolve as it moves over the aperture. However, given the prediction of aberrations over a wide field-of-view, the temporal constraints in terms of the Greenwood frequency, f_G, are further constrained due to a wider field (as compared to isoplanatic imaging) and commensurate with this, a longer period over which high-order aberrations can evolve.

The limitations of Taylor's hypothesis are of importance in the research reported in the thesis. Simulations for predicting phase perturbations at the exit pupil were conducted over field angles approaching 100 arcseconds (484 μrad). However, the practical application of this method is dependent on the validation of Taylor's frozen turbulence hypothesis over such an extremely wide field. A study using a single layer of turbulence and one reference object has been reported to establish Taylor frozen turbulent flow [37]. However, the literature is sparse on determining the validity of the Taylor hypothesis over a wide field. To address these questions, an analysis was conducted on experimental data. The results of this analysis are presented in Section 4.7.

4.3.3 Residual phase error

Residual phase error is an important metric used to measure wavefront distortion and quantify the effectiveness of removing residual phase distortion. Furthermore, AO systems can be assessed on their capability to remove wavefront phase error. In terms of this study however, theoretical phase screen variances are compared with variances taken from simulated

time series ensembles. These ensembles comprise Zernike coefficients to the 20^{th} modal order. The results provide a confidence level for wavefront propagation simulations detailed in Chapter 6 .

To determine residual phase error, wavefront phase error in the aperture, ε_W^2, is defined. In terms of modal wavefront aberrations, such as Zernike polynomials, the wavefront phase error in the aperture can be expressed, without proof, as, [117]

$$\varepsilon_W^2 = \sum_{i=1}^{\infty} \overline{a_i^2}. \tag{4.19}$$

Realistically, the accuracy of ε_W^2 in Equation 4.19 is determined by a finite number of Zernike terms. Therefore, the residual wavefront error, ε_R^2, is defined as the removal of a summed total of modal expansions, such as Zernike coefficients, from the averaged phase variance of the wavefront to be corrected. Thus, the residual wavefront error is [117]

$$\begin{aligned} \varepsilon_R^2 &= \int W(\mathbf{x}_1)\, \overline{\psi_R^2(\mathbf{x}_1)}\, \mathrm{d}\mathbf{x}_1, \\ &= \int W(\mathbf{x}_1)\, \overline{\psi^2(\mathbf{x}_1)} - \sum_{\{i_c\}} \overline{(a_i^2)}\, \mathrm{d}\mathbf{x}_1, \end{aligned} \tag{4.20}$$

where $\overline{\psi^2(\mathbf{x}_1)}$ is the averaged, wavefront phase within the aperture, and $\{i_c\}$ is the (finite) set of Zernike modes that require correction.

The covariance matrix of Zernike coefficients, Γ_a, provides a basis to calculate residual phase error

$$\Gamma_a = \overline{\mathbf{a}\mathbf{a}^T}, \tag{4.21}$$

where $\mathbf{a} \in \{a_2, a_3, \cdots, a_N\}$ and where N is the maximum number of Zernike coefficients.

By employing both the azimuthal and radial orders of the i^{th} and j^{th} Zernike polynomial, Noll formulated the lower-order covariance matrix [98], as given in Table 4.3.

From Table 4.3, Tilt phase perturbations (Zernike coefficients a_2 and a_3 shown in columns 2 and 3, respectively), comprise 86% of the total, piston removed, wavefront error, ε_W^2.

The residual phase variances, listed in Table 4.4, are shown for two turbulence profiles, $\varepsilon_R^2(10)$ and $\varepsilon_R^2(20)$. The parameter used for each profile is the Fried coherence length, r_0, and represents a comparison between turbulence strength, as a function of coherence length

Table 4.3 Covariance matrix of Zernike coefficients, $a_2 \cdots a_{10}$, normalised by $\left(\frac{D}{r_0}\right)^{-5/3}$ [117].

	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
a_2	0.448	0	0	0	0	0	-0.0141	0	0
a_3	0	0.448	0	0	0	-0.0141	0	0	0
a_4	0	0	0.0232	0	0	0	0	0	0
a_5	0	0	0	0.0232	0	0	0	0	0
a_6	0	0	0	0	0.0232	0	0	0	0
a_7	0	-0.0141	0	0	0	0.00618	0	0	0
a_8	-0.0141	0	0	0	0	0	0.00618	0	0
a_9	0	0	0	0	0	0	0	0.00618	0
a_{10}	0	0	0	0	0	0	0	0	0.00618

of the atmosphere and the diameter of the pupil. Thus, piston removed phase variances are show for each individual mode in columns 4 and 7.

To ensure propagation simulations were accurately calibrated, a time series data were collected to calculate residual phase error in the pupil. Several parameters, such as the size of the phase screen, A, the Fried parameter, r_0, and aperture, D, were required. The size of the phase screen represented the number of sample points used in this analysis. A phase screen was created using the *mid-point* method [47], where $A = 5000$, $r_0 = 10$ cm, and aperture size, $D = 1$ m.

Wavefront propagation of a single, point-source beacon was performed over the profile, $L(x,y)$. Zernike coefficients representing subsequent wavefront perturbations along $L(x,y)$ were obtained using a curvature wavefront sensor. The resulting data ensembles represented a time series of individual Zernike coefficients, $\{a_2, a_3, \cdots, a_{20}\}$. In Figure 4.4, an extract of a phase screen used for calibration of residual Zernike coefficients is shown, where the profile, $L(x,y)$, has been highlighted. The results of this simulation were used in a study to compare theoretical residual phase with the residual phase obtained from simulated wavefront data.

To determine the residual phase error of simulated phase screen data, time-series ensembles of Zernike coefficients, $a_2 \cdots a_{20}$, were constructed using M samples. A statistical analysis was performed on each ensemble. The arithmetic mean of each coefficient was calculated and compared with the theoretical, pupil averaged, mean squared phase error, $\varepsilon_R^2(\cdot)$, where

Table 4.4 Residual phase error derivations for $\varepsilon^2(10)$ and $\varepsilon^2(20)$.

Mode	$\boldsymbol{\varepsilon_R^2(D/r_0 = 10)}$	$\boldsymbol{\Delta\varepsilon^2(10)}$	Residual	$\boldsymbol{\varepsilon_R^2(D/r_0 = 20)}$	$\boldsymbol{\Delta\varepsilon^2(20)}$	Residual
1	47.8037	0.0000		151.7674	0.0000	
2	27.0140	20.7897	0.9998	85.7643	66.0031	0.9998
3	6.2197	20.7943	1.0000	19.7464	66.0178	1.0000
4	5.1522	1.0676	0.0513	16.3571	3.3893	0.0513
5	4.0846	1.0676	0.0513	12.9678	3.3893	0.0513
6	3.0077	1.0768	0.0518	9.5490	3.4188	0.0518
7	2.7246	0.2831	0.0136	8.6501	0.8989	0.0136
8	2.4368	0.2878	0.0138	7.7365	0.9136	0.0138
9	2.1491	0.2878	0.0138	6.8228	0.9136	0.0138
10	1.8613	0.2878	0.0138	5.9092	0.9136	0.0138
11	1.7499	0.1114	0.0054	5.5555	0.3537	0.0054
12	1.6338	0.1160	0.0056	5.1871	0.3684	0.0056
13	1.5224	0.1114	0.0054	4.8334	0.3537	0.0054
14	1.4110	0.1114	0.0054	4.4798	0.3537	0.0054
15	1.2950	0.1160	0.0056	4.1114	0.3684	0.0056
16	1.2393	0.0557	0.0027	3.9345	0.1768	0.0027
17	1.1836	0.0557	0.0027	3.7577	0.1768	0.0027
18	1.1279	0.0557	0.0027	3.5809	0.1768	0.0027
19	1.0768	0.0511	0.0025	3.4188	0.1621	0.0025
20	1.0211	0.0557	0.0027	3.2419	0.1768	0.0027

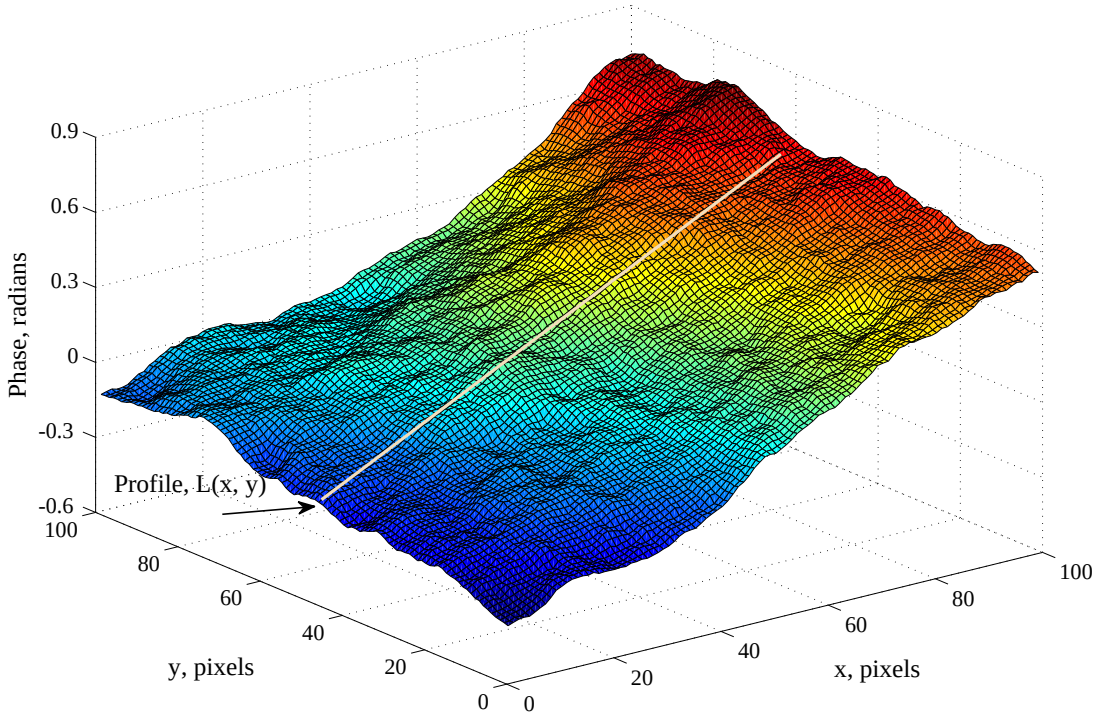

Figure 4.4 Phase screen profile used for time series statistical analysis of residual phase calculation.

$R \in \{2,3,\cdots,N\}$. The results of this comparison are shown in Figure 4.5.

In analysing the results of Figure 4.5, the mean square residual error, after removal of the first N Zernike modes, shows a dominance of tilt, a_3 Zernike coefficient, when compared to the theoretical value. This is also shown by close inspection of Figure 4.4, where a dominance of tilt aberration is evident. Irregularities such as this are a consequence of limited samples, M used in data acquisition. In this simulated analysis, a limited number of samples were used, i.e., $M = 5000$. However, using a set of randomly generated phase screens, the central limit theorem can be applied to show convergence to the theoretical residual phase error, as a function of the Fried coherence length, r_0, and aperture, D.

Lastly, to conclude this section on residual wavefront error, a commonly used metric in adaptive optics will be discussed - this metric is known as Strehl ratio, SR. The extended Marechael approximation relates the normalised intensity at the diffraction focus of an

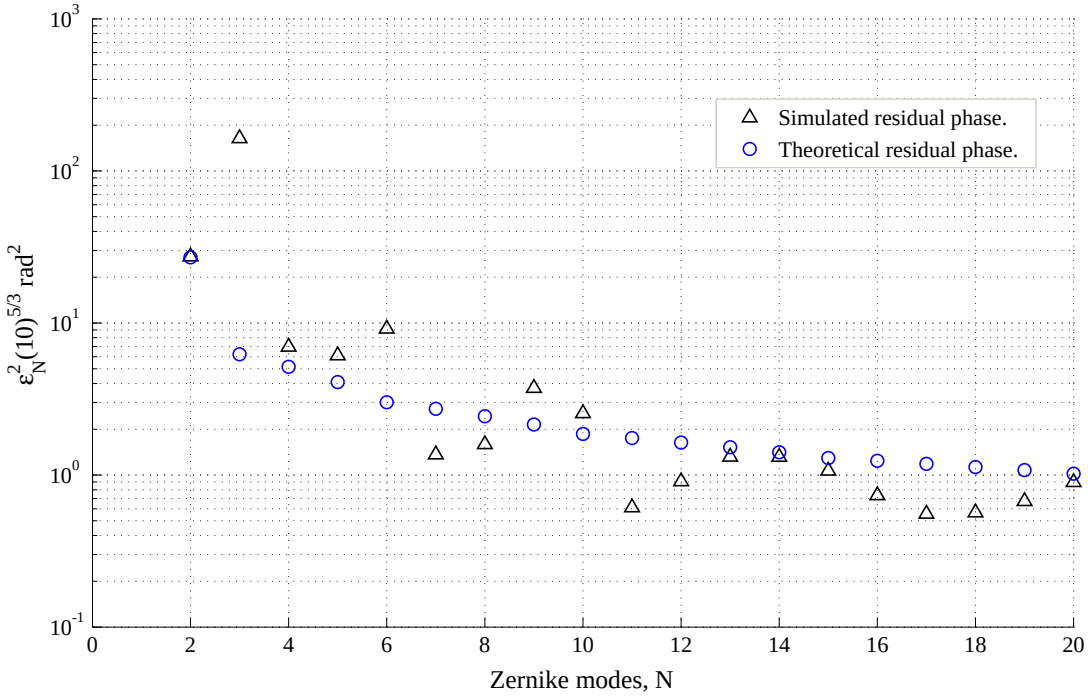

Figure 4.5 Comparison of the theoretical residual phase with simulated pupil averaged mean squared phase after removing N Zernike modes, as a function of the number of modes removed; both calculated and measured residual phase were based on $r_0 = 10$ cm, and aperture, $D = 1$ m.

optical system to the residual phase error, ε_R^2, and is given by [49]

$$\text{SR} = \exp\left[-(\varepsilon)^2\right]. \tag{4.22}$$

The Strehl ratio can also be obtained by comparing the peak of an aberrated PSF to the peak of an unaberrated, i.e., a diffraction limited PSF [117]. Given the condition of incoherent light, SR can be expressed as,

$$\text{SR} = \left.\frac{h_{ab}(x,y)}{h_{dl}(x,y)}\right|_{x=0,y=0}, \tag{4.23}$$

where $h_{ab}(\cdot)$ is an aberrated PSF, and $h_{dl}(\cdot)$ is the diffraction limited PSF.

Considering the practicalities of an optical system imaging through turbulence, a system is said to be diffraction limited if $\varepsilon^2 < 0.2$ given Equation 4.22, and SR ≥ 0.8 in terms of Equation 4.23.

4.4 Wavefront sensors

A wavefront sensor (WFS) is a device that measures wavefront aberrations caused by a perturbing medium, such as atmospheric turbulence. In this study, wavefront phase variation, rather than amplitude (scintillation), is of particular interest. Wavefront phase cannot be viewed or imaged directly; it is light intensity or irradiance, rather than phase that can be observed directly [36].

This section provides a summary of WFSs that can measure wavefront phase and describes a novel configuration that was developed for wide FOV imaging use in this study. Many of the technical terms and background topics concerning the optical wavefront are defined in Section 3.1.

4.4.1 Background

Wavefront sensors (WFSs) measure specific properties either in the optical pupil plane, or indirectly at the image plane; the former type uses incident light in the pupil of an optical system, whereas the latter employs one or more focal plane sensors to recover the wavefront.

Focal plane image sensors typically use a wavefront sensing technique known as phase diversity [39]. Two images are used: one is a focused, the second is defocused by a known amount. The wavefront is resolved using a method known as *phase diversity by analogy*, where a known phase aberration, i.e., defocus, is required to determine phase perturbations [96].

A key advantage using focal plane sensors is that additional optics are not required, other than an image sensor. However, phase diversity sensors are computationally expensive. The phase diversity technique uses an iterative search algorithm in high dimensional space to extract phase from image data [38]; phase is estimated from the OTF, $H(u,v)$, using the Gonsalves metric [39]. In order to meet real-time demands, Chundi *et al.* [22] trained an artificial neural network on a series of pre-processed discrete cosine transform (DCT) images to reduce dimensionally of the problem. Once trained, the network, when supplied with similar pre-processed images from the WFS, was able to provide wavefront phase estimates in terms of Zernike coefficients.

Pupil plane wavefront sensors require additional components to be used in the optical path,

Table 4.5 Wavefronts sensor characteristics.

Pupil Plane Sensors	**Characteristics**				
	Type	**Dim.**	**NGSs**	**#Dets.**	**Measurements**
Hartmann-Shack (H-S)	zonal	low	1	≥ 4	$\overline{\partial\phi/\partial y}$, $\overline{\partial\phi/\partial x}$
Geometric (GS)	modal	med.	1	1 or 2	$\overline{\nabla^2}$, $R_\theta(x')$
Direct curvature (CS)	modal	med.	1	1 or 2	$\overline{\nabla^2}$, $\overline{\partial\phi/\partial n}$
Extended GS & CS sensors	mixed	med.	≥ 1	1 or 2	$\overline{\nabla^2}$, $\overline{\partial\phi/\partial n}$
Image Plane Sensors	**Characteristics**				
	Type	**Dim.**	**NGSs**	**#Dets.**	**Measurements**
Phase Diversity (PD)	modal	high	≥ 1	1 or 2	$h(u,v)$, $h'(u,v)$

however the processing is generally less demanding compared to focal plane sensors. For example, a H-S sensor requires an $N \times N$ array of small lenses, referred to as a lenslet array [12], to be placed in the pupil plane of the optical system. The resulting N^2 sub-images effectively divides the incident wavefront into zones, where the wavefront slope of each sub-array, or zone, can be measured. In the case of the Curvature [114] or Geometric [136] wavefront sensors, the optical path is split into two slightly defocused images of the pupil.

Table 4.5 lists various WFSs used for adaptive optics and evaluated for this study. This table also provides a summary of WFS characteristics discussed in this section. The first column lists two, principal groups of WFSs; the characteristics of each group are summarised in the remaining columns as follows: `Type` categorises each WFS into a primary subgroup, i.e., either the pupil plane is divided into a number of *zones*, or *modal* expansions are used to decompose the wavefront into a number of distinct shapes or modes; `Dim` is the dimensionality of the wavefront sensor, in terms of the complexity of output data; `NGSs` is the capability of the WFS to concurrently estimate wavefront data from N natural guide stars or source beacons; `Dets` is the number of detector elements required for WFS operation; and lastly, `Measurements` are the quantities used to estimate wavefront phase.

Due to the wide FOV requirements specific for this research, simultaneous acquisition of wavefront phase aberrations from multiple references, as they appear in the pupil, was required. The SLODAR method for turbulence profiling employs a H-S WFSs to measure

the correlation between two sets of wavefront slope data. However, the resolution required for lenslets to detect individual slope patterns from multiple source beacons would be excessive. A novel H-S WFS has been designed to estimate simultaneous phase from up to five LGSs [77]. However, such specialised WFSs are purpose built. Lastly, the high computational demands of the phase diversity sensor limited its application to this work.

Initial attempts to use the curvature and geometric WFSs [136] for this research were limited. The literature confirmed both sensors were modal and this implied one one source reference could be used [19, 49, 114]. Essentially, zonal application of curvature based sensors had not been attempted.

However, a method of regionalising the pupil, in terms of modal expansions, was proposed by Ragazonni *et al.* [110]. This work provided a convenient method for extracting modal aberrations from each source in a regionalised, or zoned pupil plane, and this was applied to the geometric and curvature sensors. An important requirement was to remove scintillation, i.e., wavefront amplitude, and only measure wavefront phase. Both the curvature and geometric WFSs achieve this. Extensions to this requirement are given in the next section.

Regionalisation of the pupil was also applied during observation runs using the 1m telescope at MJUO. A compact data acquisition system, originally used for SCIDAR research [63], was modified and employed for this research. Configuration details on the curvature wavefront sensor used in both observational runs and laboratory tests are outlined in Appendix B.1.

4.4.2 The Curvature WFS

First proposed by Roddier [114], the curvature wavefront sensor uses two defocused images to determine the curvature of the wavefront. The intra-focal image is subtracted from the extra-focal image to provide information about phase aberrations. The curvature sensor differs from the H-S sensor in that the second derivative (curvature) of the wavefront is used, rather than the first derivative (slope). Computation to extract phase perturbations is based on Fresnel diffraction, where the intensity distribution from both defocused images are required.

The curvature sensor is defined as [115]

$$\frac{I_1(\mathbf{r}) - I_2(-\mathbf{r})}{I_1(\mathbf{r}) + I_2(-\mathbf{r})} = C\left[\frac{\partial\phi}{\partial\mathbf{n}}\left(\frac{f\mathbf{r}}{l}\right)\delta_c - \nabla^2\phi\left(\frac{f\mathbf{r}}{l}\right)\right], \tag{4.24}$$

where $I_1(\mathbf{r})$ and $I_2(\mathbf{r})$ are the defocused intra-focal and extra-focal intensity images respectively, ∇^2 is the Laplacian operator, where ϕ is the phase in the pupil, $\partial\phi/\partial\mathbf{n}$ is the outward normal derivative of the wavefront at the pupil edge, δ_c is the linear impulse distribution around the pupil edge, l is the degree of defocus from the focal point, f is the focal length, and $C = \lambda f(f-l)/2\pi l$.

Curvature WFSs have two important properties. Firstly, normalisation by $I_1(\mathbf{r}) + I_2(-\mathbf{r})$ in Equation 4.24 ensures relative insensitivity to scintillation. Secondly, the amount of defocus l from focal length f, is dependent on the size of detector with respect to Fried's parameter, r_0. Defocus is required to measure curvature and reduce the effects of Fraunhofer diffraction. In the extreme case where minimal defocus is applied, i.e. $l = 0$, the curvature sensor is limited to tilt aberrations, i.e., no curvature measurements are possible [115].

As discussed in the preceding section, a curvature and geometric wavefront sensor (WFS) were used to collect data and provided a convenient basis for acquisition of Zernike coefficients for analysis. A study of wavefront sensors by Yong [19] provided a basis for comparison. Each WFS was assessed on its ability to simultaneously convert intensity measurements to phase variances; a particular requirement was to determine individual phase measurements of multiple point-source objects in the exit pupil. Excessive defocus, such as pupil plane measurement used in SCIDAR applications, overlaps closely separated objects in the pupil plane, masking angular separation data. However, minimal defocus limits curvature measurements. Consequently, determining the optimal value for l was an important requirement to measure individual curvature and retain spatial information in the pupil.

To ensure adequate spatial resolution to derive curvature, the defocus length, l, for both intra- and extra-focal images, was calculated using Equation 4.25, where f is the focal length of the optical instrument, r_0 is the Fried coherence length, and λ is the wavelength of light [135]

$$l \leq \frac{\lambda f^2}{\lambda f + r_0^2}. \tag{4.25}$$

Effectively, the defocus length equates to Fresnel blurring and determines the resolution of the wavefront estimate [19]. Figure 4.6 shows the relationship based on the equality in Equation 4.25 between the maximum defocus value and the Fried parameter, r_0.

Field observations from the McLellan 1-m f/13.5 telescope housed at MJUO, in addition to SCIDAR data collected at the site [91], were used to calculate the length of defocus, l. The

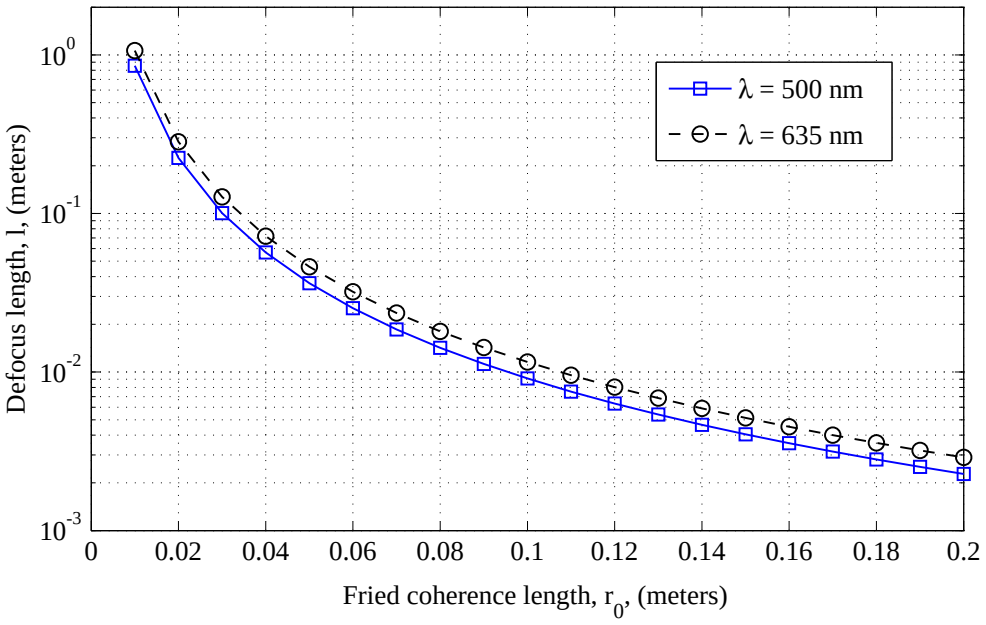

Figure 4.6 Curvature and geometric wavefront sensor defocus values, l, for laboratory and field operation.

Table 4.6 Parameters used for Curvature Sensing.

Curvature Sensor Parameters	**Values**		
	Simulation	**Laboratory**	**Field (MJUO)**
Fried Seeing Parameter, r_0 (m)	0.05 -0.2	0.05 - 0.2	0.07 - 0.18
Focal Ratio, f	13.5	13.5	13.5
Peak Wavelength, λ, 10^{-9}	500	635	500
Bandwidth, λ, 10^{-9}	400 - 600	625 - 640	400 - 600
Defocus length, l, (cm)	$3.0 \rightarrow 0.2$	$3.0 \rightarrow 0.2$	$2.0 \rightarrow 0.3$

Fried seeing parameter shown in Table 4.6 includes both the generalised and pupil-plane measurements using broadband light. The peak wavelength of the CCD spectrum was used for broadband light, however laser sources used in the laboratory were restricted to a narrow frequency band.

In summary, application of either the geometric or curvature WFSs to capture wavefront perturbations from multiple source objects required optimal values for defocus. Defocus

length for field observations ranged from 2.0 cm for severe turbulence, down to 0.3 cm for light turbulence. In the laboratory and for simulations, a slightly wider range of defocus values were used. These values are listed in Table 4.6

Performance gains, in terms of improved wavefront estimates over relatively short propagation distances, warranted investigation of the geometric WFS. An overview of this wavefront sensor is given in the following subsection.

4.4.3 The Geometric WFS

The Geometric sensor is similar to the curvature sensor, in terms of configuration and output terms. In an analysis of wavefront sensors, including the Hartmann-Shack and Pyramid wavefront sensors, the Geometric sensor was found to be a practical method of extracting modal expansions, such as Zernike polynomials, from an perturbed wavefront [19]. However, over large propagation distances performance of the geometric WFS degrades.

As described in the previous subsection, the curvature sensor uses the Fresnel kernel to convert intensity distributions, representing wavefront curvature, to modal phase measurements. However, the geometric WFS [136], as implied by it's name, applies geometrical optics to measure phase perturbations from intensity fluctuations [20].

The geometric WFS employs the same imaging system to that of the curvature sensor sensor, i.e. two defocused images are used to extract intensity profiles, however the geometric sensor uses the Radon transform to convert 2D image data into 1D *slices* whereby the conversion to phase is performed. The technique is similar to that used in computed tomography [118], such that intensity data is projected at various angles to reconstruct overall slope measurements. As with the curvature sensor, the process of converting slope measurements into modal expansions is a linear process.

In summary, the curvature and geometric sensors support different implementations, however application of the method used to image point-source objects, resulting in modal measurements of wavefront phase, is identical. The results, in terms of accuracy, were subject to certain properties and conditions specific to each sensor [19] and this provided justification to employ both sensors in this study.

4.4.4 Simultaneous wavefront measurements from multiple source beacons

In their original form, both the curvature and geometric sensors were designed to measure phase perturbations from a single source beacon. In the work presented here, the curvature and geometric wavefront sensors were extended to provide individual wavefront maps of several source objects, collectively imaged over a wide FOV; this was demonstrated in field trials and the laboratory, discussed in Section 4.7 and Appendix B, respectively. Simulation was performed using both the extended curvature and geometric WFSs and is detailed in Chapter 6. The following discussion describes the operation of these extended wavefront sensors.

As outlined in the preceding discussion on the curvature WFS, curvature is represented by disproportionate irradiance patterns on intra- and extra-focal images of the pupil. Defocus is applied so that the effects of diffraction can be disregarded, i.e. geometric optics dominates [36]. Wavefront phase can be derived from the relationship between the second derivative of the wavefront and the irradiance patterns. However, rather than measure curvature from a single source reference over the pupil, individual curvature measurements are obtained from multiple source references in the pupil.

To appreciate how multiple source measurements were achieved, consider two wavefronts ψ_1 and ψ_2 from reference source objects S_1 and S_2, separated by θ. Wavefront ψ_2 is planar, whereas ψ_1 is slightly perturbed from planar. In Figure 4.7, the intra-focal and extra-focal image planes are shown as A_2 and A_1, respectively. Irradiance patterns corresponding to, A_2' and A_1' represent the planar wavefront ψ_2 from source S_2, and A_2'' and A_1'' represent the planar wavefront ψ_1 from source S_1. Over each image plane, irradiance patterns are cropped forming two regions of interest (ROI); wavefront sensors use ROI pairs A_2' and A_1', and A_2'' and A_1'' to estimate wavefront maps of reference sources, S_1 and S_2, respectively. As stated, the wavefront from S_1 has been aberrated by turbulence, whereas wavefront S_2 remains unaberrated. Ragazzoni *et al.* [109] showed that a smaller region over the pupil can be represented by a modal subset, where values for a finite set of Zernike coefficients used to represent phase are different.

The basis of this study is to employ multiple source objects and simultaneously capture time-series data in the form of Zernike coefficients from each source for training a ESN; similar, corresponding values are then used to estimate phase perturbations from another sub-region in the pupil. In this regard, both the extended curvature and geometric sensor

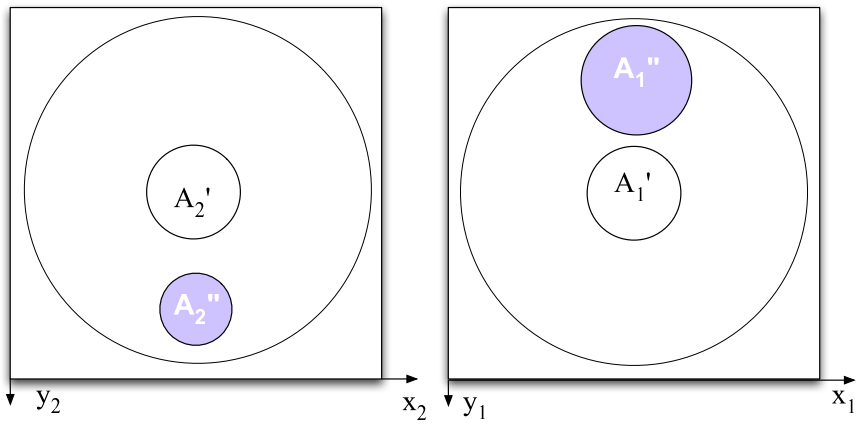

Figure 4.7 Modal representation of image planes, A_2 and A_1.

facilitated simultaneous data acquisition of Zernike coefficients from multiple reference sources. A geometric view of these extended WFSs is shown in Figure 4.8 and will be discussed in detail. A single-image WFS used to extract phase perturbations has been proposed by Hickson [52], however this was not employed due to scintillation constraints.

With reference to Figure 4.8 (a), wavefronts ψ_1 and ψ_2, supporting an angular separation of θ are incident to the entrance pupil of an optical system. The central ray of each wavefront passes through converging lens C, where rays from each wavefront converge to form a focal point. The focal length of lens C is P' and this is also the focal point for the unaberrated wavefront, ψ_2. However since wavefront ψ_1 is distorted with defocus aberration, the focal point has shifted from the focal length of lens C to point P''.

The result, in terms of irradiance distributions on defocused planes A_1 and A_2, is used to determine aberrations from the curvature of the wavefront. For example, ψ_2 is projected on both defocused planes, defined by regions $A_1'(\mathbf{r})$ and $A_2'(-\mathbf{r})$, respectively. Since the wavefront ψ_2 is unaberrated, the focal point P' is precisely at the midpoint of defocused planes A_1 and A_2, i.e., d_2' and d_1' are equivalent. This results in a corresponding irradiance profile shown by I_1' and I_2' in Figure 4.8 (b). However, P'', is not at the midpoint between defocused planes A_1 and A_2, i.e., $d_2'' < d_1''$. As a result, the irradiance profiles I_1'' and I_2'' in Figure 4.8 (b), are disproportionally distributed. A linear relationship exists between the irradiance distribution and phase perturbations that can be measured and proportioned, in terms of Zernike coefficients [19].

The application of this method can be extended to several source objects regionally dis-

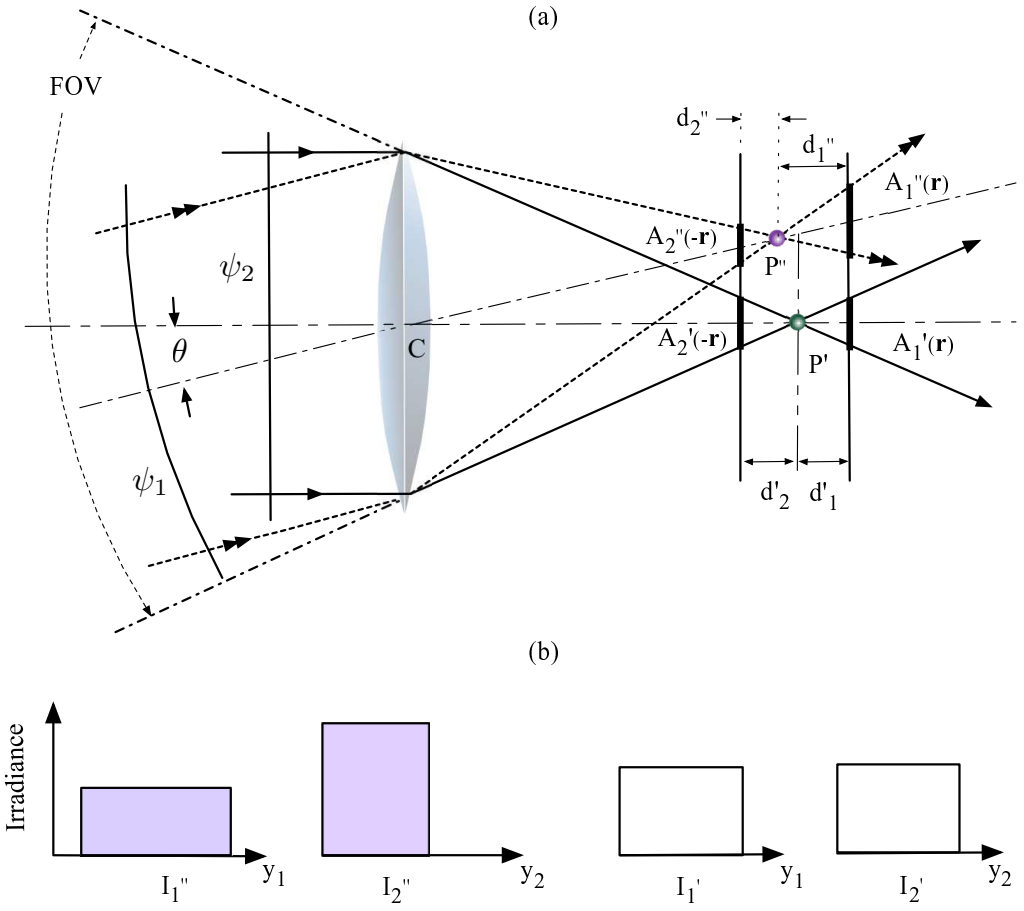

Figure 4.8 The extended curvature wavefront sensor: (a) geometrical optics view of a curvature sensor where the wavefronts from two source objects are imaged on defocused extra- and intra-focal planes, A_1 and A_2; (b) corresponding irradiance profiles for $I'_{1,2}$ and $I''_{1,2}$ of wavefronts ψ_1 and ψ_2 respectively, with angular separation, θ. Note: Due to disproportionate scaling, ROI $A''_2(\cdot)$ is shown above $A'_2(\cdot)$, however this ROI should be located in the lower portion of the intra-focal plane, A_2; the more precise orientation is shown in Figure 4.7.

tributed over the pupil and separated by various angular separations. However, as can be seen in Figure 4.8 (a), centroiding of each pair of images is required due to the offset created from the projection over wide field angles.

4.4.5 Limitations of extended curvature and geometric sensors

As discussed in the preceding subsection, sufficient defocus is required to ensure adequate spatial resolution to determine high-order aberrations. However, extending defocus to overlap either image pairs, A'_2 and A''_2, or A'_1 and A''_1, would render the extended WFS inoperable. Unlike the approach taken for atmospheric modal tomography, the ESN method described

in the next chapter does not rely on the construction of a correlation matrix of overlapping pupil data from multiple source objects. The capability to estimate the effects of turbulence, spatiotemporally, and over a wide FOV, is inherently learnt by the ESN through training.

Lastly, pixelation is directly related to inadequate spatial sampling. To limit such effects, ROIs $A'_{(\cdot)}$ and $A''_{(\cdot)}$ require regions $> 50^2$ pixels. Lastly, simulation and field studies have shown that acquisition of Zernike coefficients is limited to $< a_{11}$. This is due to a combination of two conditions. Firstly, the amplitude of Zernike modes falls steadily as the mode increases. This is due to $f^{-8/3}$ response of the Kolmogorov spectrum [49]. Secondly, the WFS is sensitive to Gaussian noise and spatial quantisation.

4.5 Deconvolution from wavefront sensing

Deconvolution from wavefront sensing (DWFS) is a hybrid system that attempts to achieve diffraction limited resolution of astronomical images from wavefront phase data and speckle (short exposure) images. This technique was first proposed by Fried [31] and was later implemented by Primot *et al.* [106]. Information about the distorting medium can be acquired using a wavefront sensor and used to calculate the PSF, **h**. The PSF, when combined with a simultaneously acquired image of a target object, **g**, can be expressed in terms of an estimate of the true object image, **f**, and noise $\boldsymbol{\eta}$ as

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \eta, \tag{4.26}$$

where **H** is commonly referred to as the distortion operator; in the spatial domain this is the PSF and combines a matrix multiplication with the operation of convolution.

DWFS can be described in terms of an open-loop AO system, i.e., a WFS is used only for image deconvolution and not for real-time corrections to the optical path of the telescope. Wavefront sensor data is employed to provide an estimate of the PSF, $\hat{h}$ and this is applied by a deconvolution algorithm on image data, **g**, to restore the original image, $\hat{\mathbf{f}}$. Tikhonov regularisation is often used in conjunction with deconvolution algorithms and can be expressed as

$$\hat{\mathbf{f}} = [\mathbf{H}^T\mathbf{H} + \beta\mathbf{C}^T\mathbf{C}]^{-1}(\mathbf{H}^T\mathbf{g}), \tag{4.27}$$

where **C** is the 2D Laplacian operator and β represents the Lagrange multiplier, also known as the regularisation parameter [8].

Conventional deconvolution algorithms, such as the Richardson Lucy algorithm [112], also known as the maximization expectation (ME) algorithm, are used for restoration. Such algorithms provide a maximum likelihood formulation, modelled on Poisson statistics. A combination of DWFS and the EM algorithm has also been proposed [23].

However, in addition to the open-loop DWFS method, Roggemann and Welsh discuss two hybrid DWFS methods [117]. Both hybrid configurations are referred to as compensated deconvolution from wavefront sensing (CDWFS) and employ an adaptive optics system to minimise the residual phase error prior to estimation of the PSF, $\hat{h}$. A *partially compensated* system supports a reduced number of wavefront sensor sub-apertures, compared with a *fully compensated* system. A CDWFS method was proposed by Roggemann *et al.* to widen the effective FOV using an off-axis guide star [116]. Leung proposed an alternative method that removed the need for a guide star through parameterisation of the transfer function [75].

Three commonly used astronomical image restoration configurations, including CDWFS and DWFS, are shown in Figure 4.9.

A configuration similar to Figure 4.9 (c) was used in this study. Justification for this approach is based on reducing the hardware complexity and cost of an adaptive optics system. Image restoration is achieved by reconstructing the SVPSF using wavefront sensor data. Three or more source beacons are employed as reference objects over a wide FOV. As a result, the incident wavefront $\mathbf{W}$ is defined as a combination of wavefront aberrations from three or more source beacons, and the DWFS module shown in Figure 4.9 (c) is replaced by an ESN, defined in the next chapter. Current methods for predicting the spatially variant PSF over a wide FOV using multiple source beacons are summarised in the following section.

In contrast to hybrid techniques, blind deconvolution [7] and the bispectrum technique [78], are exclusively post processing methods. Concerning the former, the distortion operator, i.e, the PSF, is not directly required for restoration. An initial estimate of the PSF, $\mathbf{h}_{\text{est}}$, is generated from knowledge *a priori* concerning the image; this estimate is refined using an iterative algorithm for restoration. Concerning the latter, bispectrum method, the object is reconstructed in the Fourier domain using speckle interferometry techniques and the phase component is obtained from triple correlation theory.

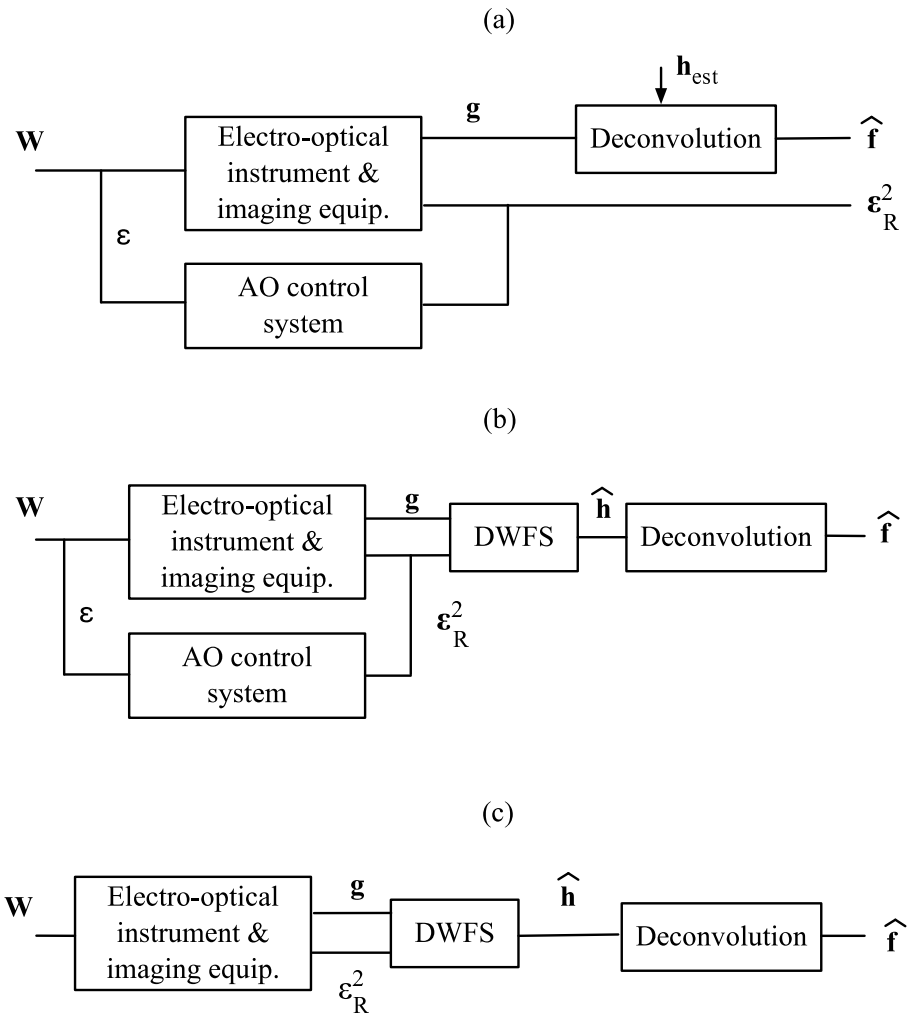

Figure 4.9 Astronomical image restoration: (a) Blind deconvolution restoration with compensated AO system; (b) CDWFS - AO compensation can be either partial or full; (c) DWFS.

4.6 Estimating the anisoplanatic PSF

Since the introduction of classical adaptive optics (CAO), a plethora of methods have been used over the last two decades to reduce the effects of anisoplanatism. Some of these methods include the use of multiple guide stars and known *a priori* turbulence statistics. Empirical data is also used to determine correlations for tomographic algorithms. These algorithms are then used to estimate the effects of turbulence on science object images, far exceeding the isoplanatic *patch* [134].

As outlined in the previous section, distorted images of astronomical science objects can be partially deconvolved using a wavefront sensor to estimate the point spread function (PSF).

Thus, only the WFS component of an AO system (operating in open loop) is required to implement DWFS. However, when the control loop is closed, i.e., wavefront phase estimates are also used to drive the AO system, further compensation to estimate the PSF for image deconvolution is achieved.

Various methods have been developed for both DWFS and compensated DWFS (CDWFS) to estimate the anisoplanatic PSF. A method that would appear effective in this regard is to employ multiple source objects to estimate the effects of turbulence over a wide FOV [126]. This method is discussed in the following subsection with two examples from the literature, modal tomography [109], and Bayesian estimation [32].

Additionally, post-processing methods are often used to improve AO images by modelling the anisoplanatic PSF. These AO-compensated methods propose to improve anisoplanatic imaging by modelling the SVPSF. The principal contributions to this field are briefly reviewed in this section.

4.6.1 Wavefront estimation using multiple source beacons

As outlined in the introduction of this chapter, AO attempts to minimise the adverse effects of air turbulence by making corrections to the optical path of telescopes. Although closed loop control is typically employed for real-time wavefront correction, in the case of DWFS, open loop operation is used to correct distorted astronomical images as part of deconvolution process. It is generally accepted that a bright natural guide star (NGS) can be used to reference a faint science object, if their angular separation is no more than of 10 μradians in the visual spectrum. However, given the low probability of such occurrences [90], one or more laser guide stars (LGSs) are often used to generate artificial references and these are placed near the science object. This is in contrast to CAO methods that traditionally use one NGS for wavefront correction [108].

Uncompensated AO systems use wavefront phase measurements from one or more source beacons to estimate the effects of air turbulence over the entire FOV. The first method that will be described was proposed by Ragazzoni *et al.*, where aberration data measured from multiple source beacons over several turbulent layers is used to improve residual phase estimates [109]. The second method by Fusco *et al.* estimates residual phase using a Bayesian approach [32]. In addition to these methods, bicoherence has been used to predict a selection of unaberrated images, referred to in the literature as *lucky regions*, over a continuous, horizontal region of turbulence [144].

Modal Tomography

With the relatively recent successes in medical tomography, Tallon and Foy [126] proposed using the individual effects on LGSs from multiple turbulent layers to form a 3D wavefront map. Given the aberration statistics *a priori* from each layer, in conjunction with the turbulent effects from multiple LGSs, covariance aberration data were used to estimate the effects of turbulence in a region that was not directly imaged. Building on this innovation, Ragazzoni *et al.* proposed a modal expansion, as opposed to a zonal approach taken by Tallon and Foy [109]. Noise reduction, and natural modal filtering around an annular pupil, were cited as justifications. The resulting method became known as atmospheric tomography [132]; however, to differentiate between zonal and modal tomography, the latter term will be used in this study.

A detailed performance analysis of atmospheric tomography is given by Tokovinin *et al.* [131] and is described with reference to Figure 4.10. Phase perturbations from three NGSs are used to estimate phase perturbations from an object, ϕ_{Obj}. A linear combination of phase measurements from NGSs are used in conjunction with Zernike decomposition to construct a linear *estimator*, also referred to as the tomographic matrix. In addition, the tomographic matrix requires a *priori* turbulence parameters. This includes the turbulence spectrum, C_n^2 profile, noise characteristics, and the heights of each layer.

In order to demonstrate the feasibility of a modal approach, Ragazzoni [109] showed that a wavefront, comprising an ensemble of Zernike polynomials of order Q, and over a large generalised pupil $\mathscr{P}^Q(x,y)$, can be described in terms of a smaller circular portion using another Zernike ensemble, $\mathscr{P}'^Q(x,y)$, inside the larger pupil. Thus, a modal approach using Zernike polynomials was demonstrated as an alternative to the original zonal method using a H-S WFS [126].

As previously mentioned, modal tomography employs knowledge *a priori* of the angular offsets of several guide stars, combined with statistical correlations from respective, overlapping wavefront maps of the generalised pupil, to estimate the effects of turbulence in anisoplanatic regions. The schematic in 4.11 shows the projection of the pupil for a target wavefront, W_T, and NGS wavefronts, S_1, S_2, S_3. This diagram can be interpreted as a *slice* through the thin turbulent layer, through which the generalised pupil is projected. Since this method requires covariance data from multiple turbulence layers, each wavefront map is marked with a subscript distinguishing the j^{th} turbulent layer.

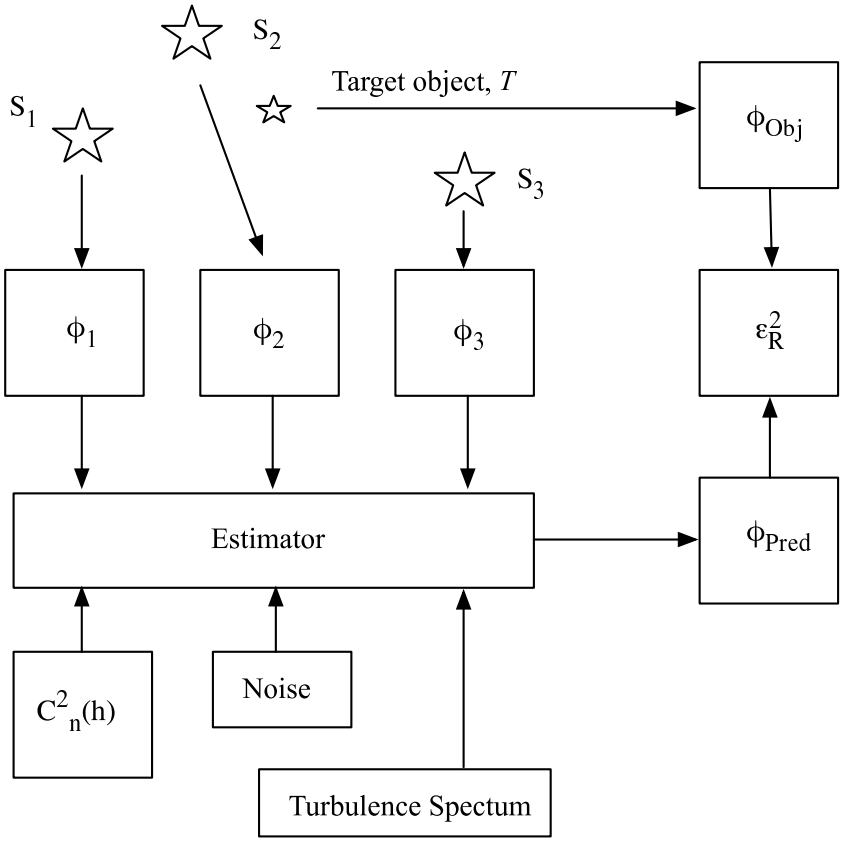

Figure 4.10 Outline of the modal tomography method. Adapted from Tokovinin *et al.* [131].

In describing this method, Ragazzoni introduces the concept of a metapupil. Over multiple layers of turbulence, the overlapping circles shown in Figure 4.11 move away from the centre. This is shown by the three arrows in the diagram. The concentric circle encompassing each pupil projection of the three NGSs and target, T_j, is shown as W_j, and is termed the *meta-pupil* [109].

For subsequent, higher altitude layers, the diameter of the meta-pupil, W_j, will expand for $j = 1 \cdots M$, where M is the total number of layers. Thus, as the overlapping portions reduce for higher altitude turbulence layers, the correlation of wavefront maps of each NGS pupil is also reduced.

As each layer is defined, a 3D tomographic profile of modal expansions is constructed using the overlapping pupil wavefront maps from multiple NGSs. As each profile is constructed, in terms of the tomographic matrix, the isoplanatic *patch* can be extended over the entire FOV.

The modal perturbation, i.e., in terms of Zernike coefficients, is defined as, a_i. For the i^{th}

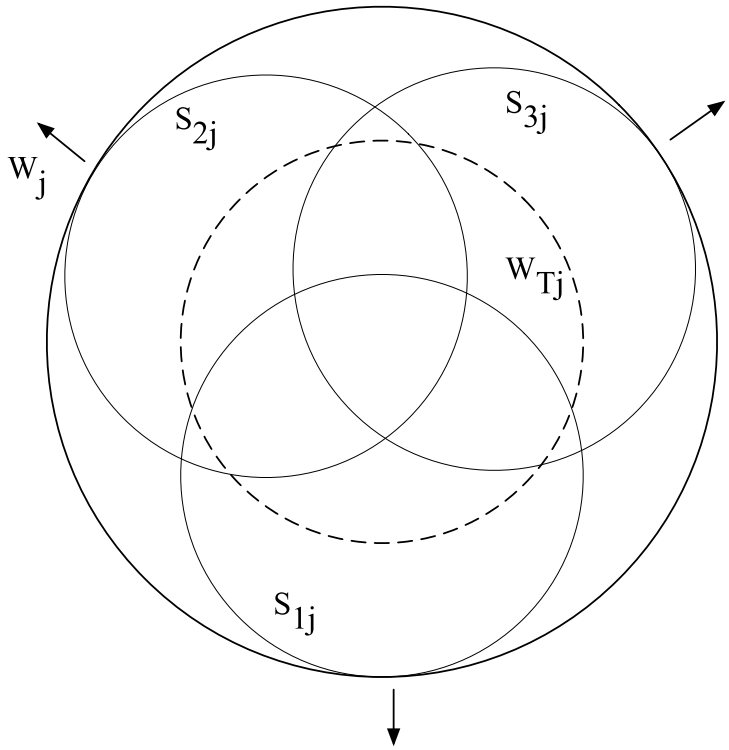

Figure 4.11 Various sections of the pupil and the metapupil, W_j [109].

NGS, and for p modal terms, S_i, can be defined as,

$$\begin{bmatrix} a_2 \\ a_3 \\ \cdots \\ a_{p+1} \end{bmatrix}. \tag{4.28}$$

Equation 4.28 can be defined in terms of a set of linear equations, for j NGSs, defining the sum of all perturbations for M layers, W_T, as [109],

$$S_i = \sum_{j=1}^{M} S_{i,j} = \sum_{j=1}^{M} A_{ij}\, W_j, \tag{4.29}$$

$$W_T = \sum_{j=1}^{M} W_{T\,j} = \sum_{j=1}^{M} T_j\, W_j, \tag{4.30}$$

where S_i is a set of NGSs $\in \{1,2,\cdots,K\}$, and where each element comprises a set of Zernike polynomials $a_2..a_{p+1}$, A is the *tomographic matrix* defining coefficients for each source over M turbulent layers, W_j is the meta-pupil of the field for M turbulent layers, and lastly, W_T is the pupil of the target object, perturbed by each turbulent layer.

Equation 4.29 can be defined in matrix form as

$$\begin{bmatrix} S_1 \\ S_2 \\ \cdots \\ S_K \end{bmatrix} = \begin{bmatrix} A_{1,1}\, A_{1,2}\, \cdots A_{1,M} \\ A_{2,1}\, A_{2,2}\, \cdots\, A_{2,M} \\ \cdots\ \cdots\ \cdots\ \cdots \\ A_{K,1}\, A_{K,2}\, \cdots A_{K,M} \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \cdots \\ W_K \end{bmatrix}, \tag{4.31}$$

The matrix shown in Equation 4.31 can be written in vector form as

$$\mathbf{S} = \mathbf{A}\,\mathbf{W}, \tag{4.32}$$

and solved for $\mathbf{W}$ using singular value decomposition [109].

One of the principal aims of this research is to replace the tomographic matrix, A, with an echo state network (ESN) for the prediction of modal perturbations, i.e., Zernike coefficients a_2 to a_{20}, within any localised region defined by the metapupil, W_j. To replace A, an ESN is trained using a new spatiotemporal algorithm and a series of known *a priori* modal aberrations within W_j. However, the overall structure shown in Figure 4.11 was used as a basis for the simulation model employed in this dissertation.

Bayesian estimation

Maximum *a posteriori* (MAP) is a linear technique that uses prior knowledge concerning a problem, in conjunction with Bayes' theorem, to maximise the most probable solution. For example, if prior probabilities such as noise statistics are known, *a priori* knowledge can be incorporated with Bayesian inference to estimate the most likely value, representing the phase within an anisoplanatic region.

To minimise the effects of anisoplanatism, phase estimates over a large FOV have been demonstrated using the MAP technique [32]. The approach requires measured C_n^2 profiles of two or more turbulent layers from several NGSs; in terms of data acquisition, this method is similar to Ragazzoni *et al.* [109]. However, rather than employ a tomographic approach,

phase variations in the pupil are estimated using *a priori* atmospheric profiles, i.e., $C_n^2(h)$, from the j^{th} turbulent layer, at height, h_j.

MAP estimators tend to be used in systems where Gaussian statistics dominate, whereas maximum likelihood estimators (ML) are prevalent in situations that can be described by Poisson statistics [75]. This would suggest that MAP estimators are perhaps not the optimal choice as wavefront estimators, especially considering photon starved applications where Poisson statistics dominate. However, Hunt *et al.* [54] propose two multiframe Poisson MAP algorithms for the deconvolution of astronomical images.

4.6.2 The AO-compensated anisoplanatic PSF

The residual phase error resulting from compensated AO systems can be further reduced using models and characterisations of both the OTF and anisoplanatic PSF. Such reductions have been applied over the entire FOV resulting in improved estimates of the anisoplanatic PSF. Two methods have been used: the first proposed by Fusco *et al.* uses angular separation measurements *a posteriori*, in terms of the Zernike coefficient correlation matrix, to characterise the optical transfer function and corresponding anisoplanatic PSF [33]. A second method proposed by Aubailly *et al.* [5] firstly models the PSF using a known source reference, then interpolates the coefficients of the model to predict the spatially variant PSF. A comprehensive overview of current methods used to estimate the AO-compensated anisoplanatic PSF is given by Britton [13].

Characterisation of the anisoplanatic point spread function

To estimate residual phase errors in anisoplanatic imaging, the spatially variant PSF can be characterised in terms of an angular correlation matrix [33]. The forward spatiotemporal image model outlined in Subsection 3.7.1 is used in this method.

Parameterisation of the anisoplanatic PSF

Aubailly *et al.* proposed a method using a parameterised model to predict the AO-corrected, long-exposure (LE) anisoplanatic PSF [5]. As distinct to other approaches outlined in this section, this method assumes certain properties characteristic of the AO-corrected PSF as it explores anisoplanatic regions over image space. For example, the LE, AO-uncorrected PSF has rotational symmetry, however due to anisoplanatism, the AO-corrected PSF does not. Such properties of the LE, AO-corrected PSF are used to construct a parameterised

model of the spatially variant PSF [5].

4.7 Qualifying wavefront deformation in the metapupil

Widening the field angle improves the probability of encompassing multiple, bright NGSs for anisoplanatic wavefront prediction [90], however for their use in anisoplanatically distorted images, other practical considerations need to be addressed. These considerations, with specific reference to astronomical tomography and the ESN method detailed in this thesis, are outlined in this section.

The method using multiple guide stars to estimate the spatially variant PSF was outlined in the previous section. Ragazzoni *et al.* [110] verified this theory experimentally. Three NGSs in the constellation of Aquila, selected for their near equilateral triangular configuration, were used to estimate wavefront aberrations of a fourth NGS at the centre. The 3.6-m Telescopio Nazionale Galileo (TNG) used a single, defocused image of the pupil to extract wavefront curvature [52]. Four wavefront maps were retrieved, corresponding to phase aberrations from each source beacon and a target. The three NGSs used as source beacons varied in magnitude, $9 \leq m_v \leq 11$, and the target, $m_v \approx 12$. Angular separations from each source to the target varied between 58.1 μrad, 76.1 μrad, and 90.7 μrad.

To test and verify the ESN method experimentally, a stellar configuration similar to that used by Ragazzoni *et al.* [110] was required. However, considering the difference in light gathering capabilities between the TNG and the McLellan 1-m telescope, suitable source beacons and target objects were limited to NGSs of magnitude $m_v \leq 6$. As a result, the configuration in the constellation of Aquila was discounted. Alternative sources however, such as Trapezium in Orion, where angular separations between the four brightest NGSs varied from 31 μrad to 106 μrad, were considered. However, high noise floors from sub-images of faint, $m_v \approx 6$ NGSs, were noted. Specifically, these included θ^1 Ori A, B, and D. Given the requirement for short exposure periods to minimise temporal decorrelation of the atmosphere, these results proved unreliable due to photon starvation. Therefore, an alternative to the use of NGSs was investigated.

Laser guide stars (LGSs) are a practical alternative to counter the sparsity of bright, natural source beacons, however their improvement in terms of optical resolution is limited by their upward propagation path through turbulence. For this research, a suitable alternative to NGSs was found in the use of planetary moons.

The four largest moons of Jupiter were considered suitable for source beacons, and their defocused images were captured over a wide FOV. Planetary moons are not point sources because of their size, however the way the light is generated is not a fundamental difference in this instance. As a result, these extended objects were considered suitable for this analysis. The planet Jupiter was occluded from the FOV to prevent reflected planetary light saturating the image. Phase perturbations from various two and three Jovian moon configurations were recorded over several data runs. As these moons orbited Jupiter, their angular separation varied over time. Such separations can be accurately measured [97], however due to the near planar orbit of the Jovian satellites, a trapezoidal configuration of four moons is a rare occurrence.

An example of defocused pupil images of moons Io and Europa is shown in Figure 4.12. Subfigures (a) and (b) are intra-focal and extra-focal images, respectively. Both subfigures have been inverted to enhance image details and show noise components. The latter was significantly reduced after post-processing each image with a large ensemble of *dark* frames [11, 53] recorded immediately after each observation run.

Given spatial constraints, in terms of parsimonious configurations, such that field angles of at least 3 moons supported $\theta_{\text{Sep}} \leq 49$ μrad, wavefront maps from the four "visible" Jovian moons were not used for ESN training and testing. However, the resulting data ensembles were used to test Taylor's hypothesis over wide field angles. Quantifying Taylor's hypothesis over a wide field angles is an important requirement for application of the ESN method. Additionally, this result may also have implications on the experimental study of atmospheric tomography. To test this hypothesis, several observational runs were conducted over various field angles. An analysis of the resulting data is presented in Subsection 4.7.3; details of each acquired data ensemble are summarised in Tables C.2 and C.3.

4.7.1 Near ground turbulence effects on anisoplanatic imaging

A series of observation runs were scheduled using the 1-m McLellan telescope at MJUO. To study the effects of turbulence on multiple source beacons, a simplified, single-layer turbulence model was required. Applying a single-layer turbulence model to field observations required a *divide and conquer* approach. However, by employing the equipment described in Appendix B.1, it was not possible to completely isolate the effects of individual layers. Using data ensembles from a series of observation runs, analysis of the resulting wavefront maps showed the combined turbulent effects of at least two layers.

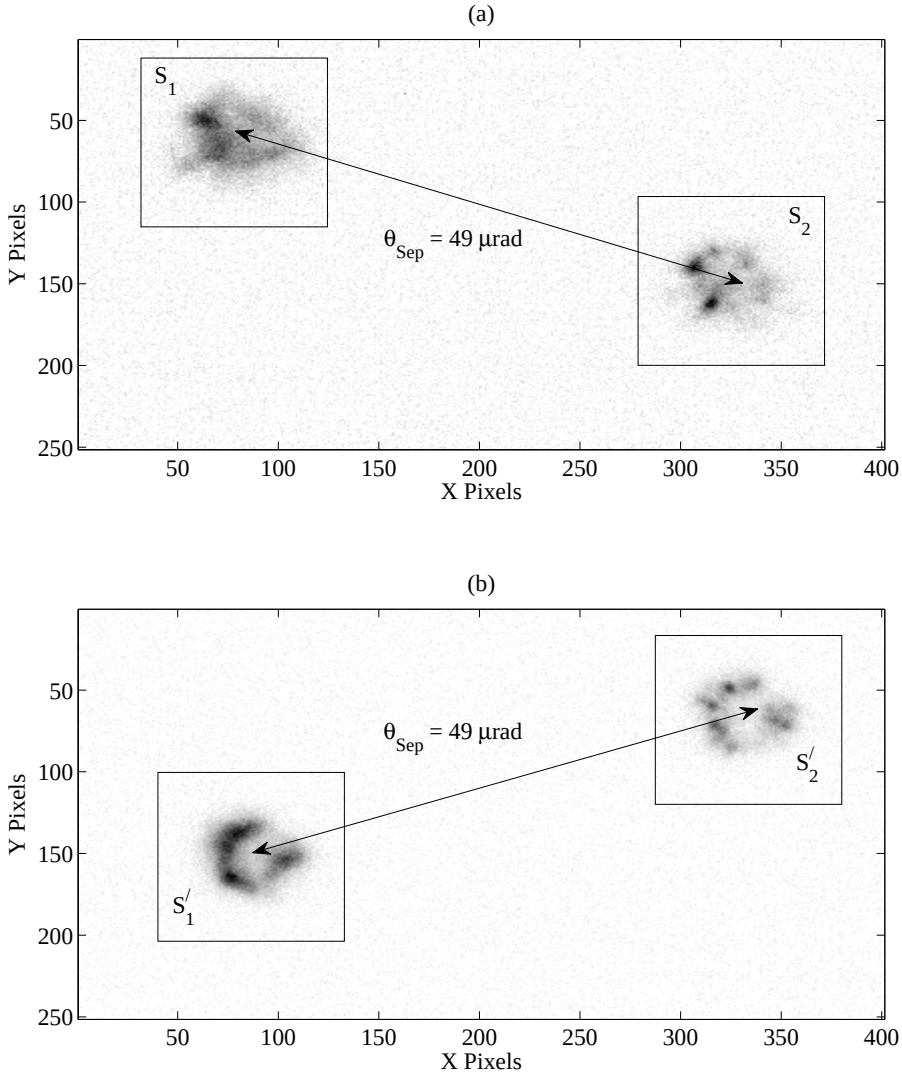

Figure 4.12 Inverted, defocused images of two Jovian moons Io and Europa, used for wavefront phase experiments: (a) intra-focal image of Io (S_1) and Europa (S_2); (b): extra-focal image of Io (S_1') and Europa (S_2') from dataset `JUP_2a27i` in Table C.2.

Fortunately, the results from a research study on atmospheric trending at MJUO described several turbulence models; each model was related to the prevailing weather and seeing conditions [91]. Two of these models, specifically, `MJUO1V` and `MJUO2V`, were employed to facilitate the isolation of a single layer of turbulence. Subsequent acquisition of wavefront data were then used for time-series analysis of individual aberrations, each pertaining to the dominant effects of a single, turbulent layer. A summary of the parameters used for MJUO models is listed in Appendix Table E.1.

The results from an earlier study by Mohr *et al.* [92] confirmed the presence of dominant near-ground turbulence (NGT). Additionally, a weaker tropospheric turbulent layer at an altitude of between 11-14 km above sea level, and a second, weak layer, detected at 6-7 km above sea level, were reported [92]. Using a set of comprehensive trending models for MJUO [91], a geometrical view of the metapupil [109] for the 1-m class telescope at MJUO was used to study the combined effect of NGT and a single, tropospheric layer.

Figure 4.13 shows a NGT layer, L_1, and a higher altitude layer, L_2. The curvature WFS discussed in Section 4.4 was used to determine individual phase perturbations from defocused images of NGSs, labelled S_1 and S_2 in Figure 4.13, and over a FOV of 480 μrad. By adopting the thin-layer turbulence model, metapupils M_1 and M_2, can be defined for turbulence layers L_1 and L_2, respectively. Based on a somewhat unrealistic scenario that both layers L_1 and L_2 move at a similar, constant velocity, i.e., $V_{\text{wind}}(h_1) = V_{\text{wind}}(h_2)$, it can be shown that the individual temporal effects on NGSs S_1 and S_2 are dominated by turbulent layer, L_1.

Formulation of this model was beneficial for two reasons. Firstly, the characteristics of layer movement and interaction were used for the propagation model detailed in Chapter 6. Secondly, this model is used in the analysis on Taylor's hypothesis in Subsection 4.7.4.

4.7.2 Wavefront sensing constraints on near ground turbulence

In this subsection, the model shown in Figure 4.13 is used to define a minimum height constraint on a near ground turbulent layer for wavefront sensing. This was necessary in order to qualify wind and seeing conditions that would otherwise render observational results unsuitable for analysis. This constraint was also incorporated in the simulation model discussed in Chapter 6.

Given a telescope of aperture A, and assuming the thin-layer turbulence model and Taylor hypothesis [128], phase perturbation $\psi(\mathbf{r}_2)$ from layer L_2, moving at a constant velocity, will result in a speckled image of S_1 at time, t_1. As the turbulent layer L_2 is moved over the FOV by prevailing winds at a velocity of $V_{\text{wind}}(h_2)$ and in the direction of S_2, wavefront $\psi(\mathbf{r}_2)$ is displaced by distance, d_2; as a consequence, and still within the FOV, a corresponding speckle will effect S_2, $t_4 - t_1$ seconds later.

Now consider similar conditions for layer L_1. Assuming the same constant velocity of turbulent layers L_1 and L_2, the wavefront $\psi(\mathbf{r}_1)$ will be displaced by d_1 and take only $t_3 - t_2$ seconds before the speckled image of S_1 is repeated for S_2. Thus, in terms of temporal

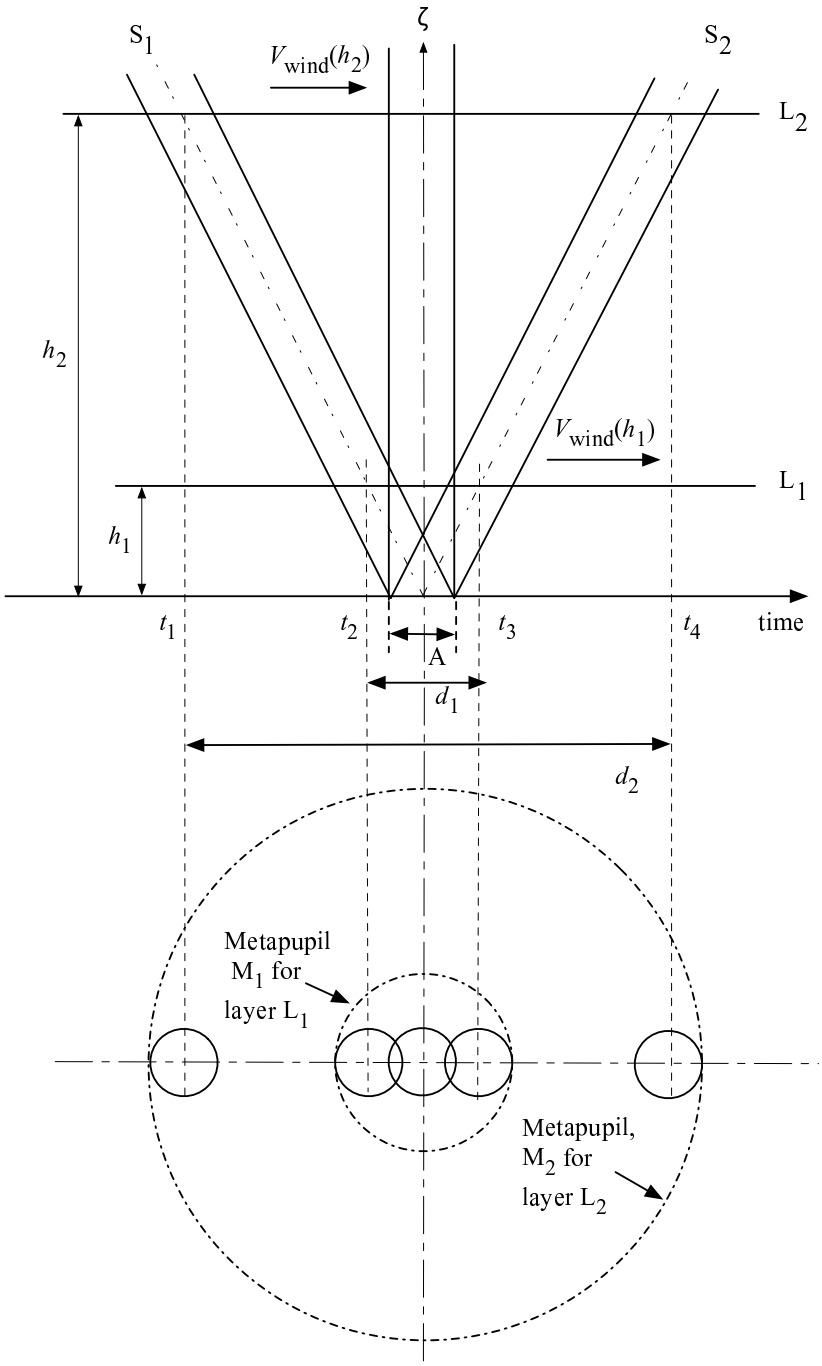

Figure 4.13 Geometrical view showing the effect of a dominant low-altitude boundary layer.

decorrelation, and given the (highly) relaxed constraint that a similar structure constant applies to both layers, i.e., $C_n^2(h_2) \approx C_n^2(h_1)$, the effectiveness of capturing individual perturbations resulting from speckle images of NGSs over a single layer is increased if the dominance of the NGT layer is reduced.

Using optimal or known *a priori* values for parameters such as sample rate, f_S, velocity of the turbulent layer, $V_{\text{wind}}(h)$, and field separation angle, θ_{FOV}, the minimum height of a turbulent layer, detectable using the equipment outlined in Appendix B.1, and that can be applied to the ESN method, is

$$h_{min} = \frac{V_{\text{wind}}(h)}{f_S} \cdot \arctan\left(\frac{\theta_{\text{FOV}}}{2}\right), \tag{4.33}$$

where $V_{\text{wind}}(h)$ is given in $\mathrm{m\,s^{-1}}$.

Given the dominance of NGL turbulence, calm conditions with good seeing are required to study the effects of the tropospheric turbulence layer. Under such prevailing conditions we can assume, given the `MJUO1V` and `MJUO2V` models previously introduced, that the tropospheric turbulence layer is moving at a velocity, $V_{\text{wind}}(h)$, of either 12 $\mathrm{m\,s^{-1}}$or 30 $\mathrm{m\,s^{-1}}$, respectively [91]. Assuming the latter, Equation 4.33 is used to calculate the minimum height of non-boundary layer turbulence, h_2, such that, $h_2 = 1.25$ km. Given that MJUO is at an altitude of 1027 m, the minimum altitude of turbulent layer L_1, for practical application of the ESN method, is ≈ 2.28 km above sea level.

Based on the above discussion, and in consideration of MJUO turbulence models [91, 147], field conditions and equipment parameters have been established for acquisition of observational data for time-series analysis of wavefront phase data, and ESN training and testing. In summary, ideal conditions are required to minimise the effects of a strong boundary layer and to capture the effects of a single, upper tropospheric layer at a height, according the MJUO model, of 11 km.

4.7.3 Review of Taylor's frozen hypothesis

The Taylor hypothesis described in Subsection 4.3.2 is used as the basis for simulations and observational results in this work. Due to its influence on the ESN method, the validity of Taylor's frozen turbulence hypothesis was investigated, specifically, with application over a wide FOV. According to Chun [21] the useful FOV for an adaptive optics system is dependent on the wavelength, and the amount of correction obtained must be sacrificed as the isoplanicity decreases for $\theta \geq 290\,\mu$rad. The ESN method can be applied over this field angle, however the accuracy of estimations is constrained by decorrelation of the atmosphere. Thus, the importance of quantifying Taylor's hypothesis over a wide FOV.

In an attempt to validate Taylor's hypothesis, Gendron *et al.* calculated the spatiotemporal

cross-correlation function of wavefront slopes using a H-S WFS [37]. Motivation for their research was the suggestion that *frozen phase slabs* could hold potential advantages, in terms of improving AO efficiency. Cross-correlation was performed on two sets of H-S slope measurements, given by the spatiotemporal expression [37]

$$R_{xyt} = \frac{\left\{ \int d\zeta \int d\eta \mathbf{S}(\zeta,\eta,\tau) \cdot \mathbf{S}(\zeta+x,\eta+y,\tau+t) \right\}_{\tau}}{\int d\zeta \int d\eta \mathscr{P}(\zeta,\eta) \cdot \mathscr{P}(\zeta+x,\eta+y)}, \tag{4.34}$$

where $\mathbf{S}(\cdot)$ is the set of slopes from the HS-WFS, $\mathscr{P}(\cdot)$ is the generalised pupil function and ζ, η, and τ are spatiotemporal parameters in the object plane, and x, y, and t are spatiotemporal parameters in the image plane.

With the application of Equation 4.34, Gendron *et al.* [37] used the normalised maximum correlation peak, and subsequent peaks, to calculate the results given in Table 4.7.

Table 4.7 Correlation peak maximums, normalised to $R(\tau = 0)$ [37].

τ (ms)	0	40	80	120	160	200	240	280
Correlation	1.00	0.69	0.59	0.53	0.51	0.49	$\cdots$	$\cdots$

From temporal measurements in the focal plane, these results confirmed decorrelation of the atmosphere over the first 80 ms from a wavefront first being detection over the aperture. However, over the remaining periods, subsequent decorrelation reduced considerably. The authors suggest that this reduction is an example of a "sliding phase slab" passing over the aperture, and evidence of Taylor's hypothesis. However, given a wider FOV and/or lower wind velocity, $V_{\text{wind}}(h)$, the rate of decorrelation may have increased before the wavefront was occluded by the adjacent edge of the exit pupil. Therefore, further consideration was given to these effects.

According to McGuire *et al.* [90], each turbulent layer can be considered *frozen*, since the timescales for the turbulence to change are of much longer duration than it takes for the turbulence to be blown across the aperture. However, as outlined by Gendron *et al.* [37], wavefront aberrations, such as tilt, vary due to temporal decorrelation of the atmosphere. As discussed in the preceding section, and with reference to Figure 4.13, several configurations and conditions are responsible for deformation of the wavefront while it is measurable in the metapupil. These can be summarised as:

- Low wind velocity, $V_{\text{wind}}(h)$.
- Dominant, near ground layer (NGL) turbulence.
- A wide field-of-view (FOV).
- Additional turbulent layers with varying wind velocities.

For example, let the metapupil for layer j, be represented by W_j. For the following analysis, the Greenwood frequency, given by Equation 4.11, is required. However a simplified expression, assuming observations at zenith and a constant wind velocity, is given by Tyson [134]

$$f_G = 0.43\left(\frac{V_{\text{wind}}(h)}{r_0}\right), \tag{4.35}$$

where r_0 is the Fried coherence length in meters, and , $V_{\text{wind}}(h)$ is a constant wind velocity at height, h, measured in $\mathrm{m\,s^{-1}}$.

To maintain the coherence of the atmosphere, given an average wind velocity of $21\,\mathrm{m\,s^{-1}}$ [91] and $r_0 = 0.12$ m, the Greenwood frequency for MJUO is ≈ 75 Hz. With good seeing and calm conditions, other MJUO models can be used, supporting more relaxed requirements [91]. However, to quantify Taylor's hypothesis, in terms of the above criteria and in accordance with MJUO models discussed in this section, an experiment was devised. The procedure, and subsequent analysis of the results from this experiment are detailed in the remainder of this section.

4.7.4 Quantifying Taylor's hypothesis for wide-field imaging

Given the experimental configuration discussed in the preceding sections, a method was devised to test the validity of Taylor's hypothesis over wide field angles. Firstly, the extended geometric sensor was used to measure wavefront phase simultaneously from two source beacons. Since the Jovian moons method outlined earlier in this section was used in this study, the angular separation of each beacon varied predictably over time. Secondly, time series data were acquired at a rate exceeding the Greenwood frequency, f_G; temporal phase differences in these measured data were compensated by time shifting one of the series by a period commensurate with the estimated wind velocity. Lastly, a correlation coefficient matrix comprising Zernike coefficients was constructed. This provided independent data analysis on each time series and a collective comparison was conducted on empirical data ensembles.

The effects of turbulence were measured over various field angles using three Jovian moons, Io, Europa, and Ganymede, between 11:33 and 16:12 GMT, on 28 July 2009. Calm, clear conditions, with excellent seeing and a light Easterly wind, $V_{\text{wind}}(0) \approx 2\,\text{m}\,\text{s}^{-1}$, prevailed. The apparent magnitudes from the Jovian moons used in this study were relatively consistent, where $m_{v_1} \approx 5.02$, $m_{v_2} \approx 5.29$, and $m_{v_3} \approx 4.61$, respectively; accurate trajectory and angular separation data were provided by the Almanac [97]. Details of observation runs to acquire data for this analysis are provided in Appendix C.

To ensure correlation was performed on the same, time-delayed wavefront data, a temporal shift is applied to the leeward data ensemble, w.r.t. to the windward ensemble. Wind velocity profiles, `MJUO1V` and `MJUO2V`, were initially used to correlate data from measured wavefront data [91]. According to Mohr [91], a tropospheric layer at an altitude of 11 km travels at a wind speed of 12 $\text{m}\,\text{s}^{-1}$and 30 $\text{m}\,\text{s}^{-1}$, respectively. Based on these parameters, time series data were temporally shifted by N_S samples, where

$$N_S = \frac{\Delta T_m}{\Delta T}, \tag{4.36}$$

and where ΔT_m is the period, based on a wind model for the wavefront to be displaced in the metapupil, and ΔT is the sample period of the recording system.

By applying wind velocities V_1 and V_2 from wind models `MJUO1V` and `MJUO2V`, respectively, and using a sampling period of $\Delta T = 16.67$ ms, Table 4.8 shows the theoretical number of samples required for temporal phase compensation for the leeward Zernike ensemble over an angular separation, θ_{Sep}.

Table 4.8 Temporal phase shift parameter summary, calculated for a dominant, single layer of turbulence at a height of 11 km and velocities, $V_1 = 12\,\text{m}\,\text{s}^{-1}$and $V_2 = 30\,\text{m}\,\text{s}^{-1}$.

Ensemble	Frames (10^3)	θ_{Sep} (μrad)	d_{L_H} (m)	ΔT_1 (V_1)	ΔT_2 (V_2)	N_1 (V_1)	N_2 (V_2)	Source beacon[a]
`JUP_2a28h`	5	320	3.525	0.293	0.117	+17	+07	I, II
`JUP_2a28i`	5	560	6.163	0.513	0.205	+30	+12	II, III
`JUP_2a28a`	10	730	8.034	0.669	0.268	-40	-16	II, III

[a] Io = I; Europa = II; Ganymede = III.

Data ensembles from wide field observations of two Jovian moons were collected over

several observation runs at MJUO. Two groups of data are shown in Appendix Tables C.2 and C.3. The second group was acquired under excellent seeing conditions, and was used in this analysis. Three time-series ensembles, each comprising two subsets of aberration data, i.e., one set for each moon, were individually processed; each set is composed of 19 Zernike coefficients. A correlation analysis was performed on each set from the three ensembles, `JUP_2a28a`, `JUP_2a28h`, and `JUP_2a28i`. Given the conditions outlined in Subsection 4.7.2, the results of this analysis are shown in Figure 4.14.

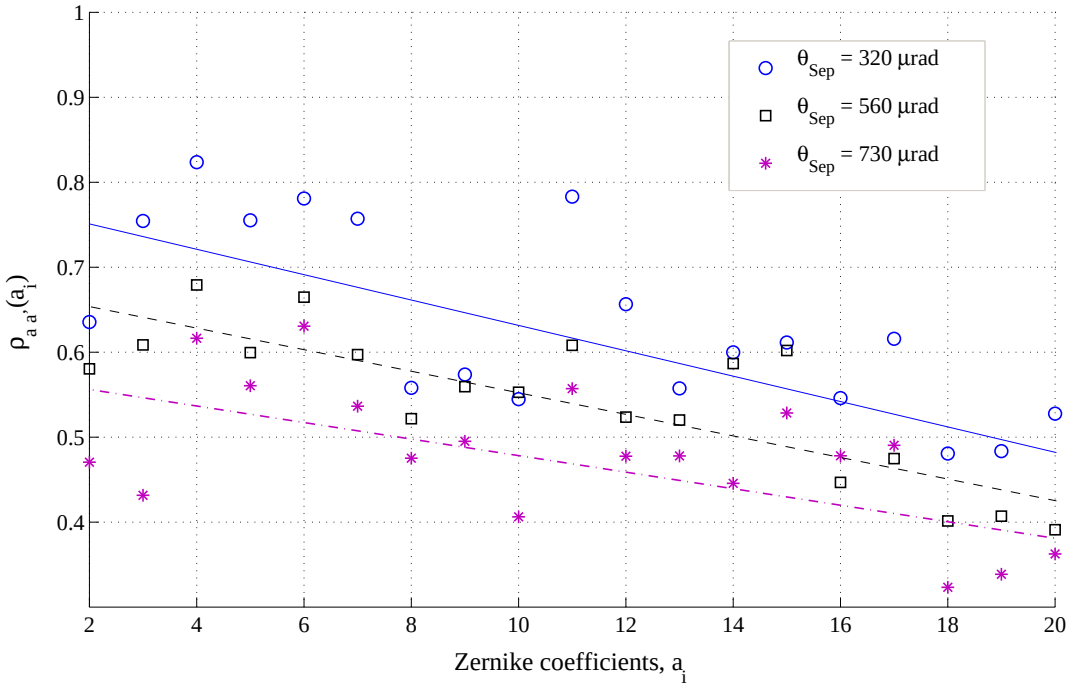

Figure 4.14 Correlation of Zernike coefficient data set, $\mathbf{a} = \{a_2, \cdots, a_{20}\}$, acquired from three angular separations, $\theta_{\text{Sep}} \in \{320, 560, 730\}\,\mu$rad, between two Jovian moons.

Pearson's correlation coefficient $\rho_{\mathbf{a},\mathbf{a}'}(\cdot)$ was calculated from two time series ensembles. These ensembles are composed of a set of Zernike coefficients, $\mathbf{a}$ and $\mathbf{a}'$, where $\mathbf{a}$ and $\mathbf{a}' \in \{a_2, \cdots, a_{20}\}$. Each Zernike ensemble represented 1000 wavefront maps of time-series image data for each moon. Pearson's correlation coefficient, ρ_{XY}, is given as

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sigma_X\,\sigma_Y}, \tag{4.37}$$

where σ_X and σ_Y are the standard deviations of X and Y, and $\text{cov}(X,Y)$ is the covariance of ensembles X and Y, and is valid for, $-1 \leq \rho_{XY} \leq 1$.

Close inspection of Figure 4.14 shows distinct groups of three distinct groups of data, where each group represents an ordering of Zernike coefficients, $a_2, \cdots, a_{20}$. The group with the highest correlation had the shortest separation, i.e., 320 μrad. This is followed by the next widest separation, 560 μrad, and lastly, 730 μrad. To highlight trending of each set of Zernike coefficients, linear regression was performed on each ensemble; this is represented in Figure 4.14.

Figure 4.14 demonstrates how 'frozen' turbulence can vary over a wide field angles. It also suggests that high-order Zernike coefficients are less correlated than low-order terms. This was expected, since higher order aberrations contain less energy and momentum, thus susceptible to more chaotic behaviour. Notable exceptions in this analysis are the tilt terms, a_2 (x-component), and a_3 (y-component). Since these terms are more sensitive to WFS errors due to centroid misalignments, further data pre-processing may correct such inconsistencies.

A similar analysis was performed using data acquired over smaller angular separations, i.e., $50 \leq \theta_{\text{Sep}} \leq 300$ μrad. However, changeable weather and variable seeing conditions during acquisition warranted a different approach. Based on the consistency of results presented in Figure 4.14, and other comparisons using data ensembles listed in Appendix C, the limits of Taylor's hypothesis were qualified in terms of field angle, and with respect to the height of a single turbulent layer. Application of the ESN method for SVPSF prediction should also be governed by the same constraints.

Lastly, the results shown in Figure 4.14 were extrapolated to provide a field limit for application of the ESN method. As stated earlier in this section, field angles for closed loop (AO) systems can exceed 200 μrad, however this is at the expense of correction. For open loop systems, such as DWFS, tighter constraints are expected. Therefore, angular separations between a source beacon and target object are likely to be adequately constrained, according to the Taylor hypothesis for a single layer of turbulence at a height of 11 km, over the range, $0 \leq \theta_{\text{Sep}} \leq 120$ μrad.

Chapter 5

Reservoir Computing

Artificial neural networks (ANNs) are inspired by biological neurons and their interconnections that form the structure of the brain. ANNs can be defined as a network of nodes, each comprised of one or more inputs and outputs, a matrix of weighted interconnections providing communication between each input and output, and an activation function ensuring each output is bounded [51]. When an analytical model is either not available or is too complex, *black-box* modelling can be applied using ANNs to find a solution. In the work presented here, ANNs do not necessarily model internal mechanisms, but are trained to learn the overall behaviour of a target system [59].

One class of ANNs, generically referred to as recurrent neural networks (RNNs), support an architecture based on signal feedback. Such networks can be used to enhance a system's ability to predict some future value, or to capture the dynamics of input signals [85]. The integration of RNNs and associated topologies for control applications has augmented or replaced many existing models by providing a simplified methodology. A specialised family of recurrent neural networks, referred to in the literature as *reservoir computing* [121], is introduced in this chapter. More specifically, the significance of the echo state network (ESN), recently promoted as a tool to solve many engineering problems, is discussed.

The organisation of this chapter is as follows. A brief background on learning methods is given in the next section. This is followed by a discussion on design considerations and an introduction to reservoir computing in Section 5.2. The echo state network (ESN) is introduced in Section 5.3, where a generic approach is taken to define the architecture. Training methodologies are considered and the principal data structures used for spatiotemporal training are defined. Key parameters used in the design of echo state networks for

single time-step prediction are also defined in Section 5.4, and these parameters are used for system optimisation. A performance analysis, based on design parameters used for this implementation, is also conducted. Considerations specific to this application, such as noise and memory issues, are outlined in Section 5.5. Lastly, two fundamental terms used throughout the reminder of this dissertation are qualified in Section 5.6.

5.1 Background

Traditionally, ANNs have been used in signal processing for classification, auto-association, and prediction [87]. In the field of adaptive optics, ANNs have been used in two key areas - classification and prediction. Firstly, wavefront aberrations measured as irradiance patterns in the focal plane have been classified in terms of Zernike polynomials. Two ANN architectures used to achieve this include radial basis functions and multi-layer perceptions [22]. Such ANN implementations have also been used to control the output of an adaptive mirror [137]. Secondly, prediction of wavefront aberrations has been used to partially compensate AO systems for computational latency [34, 66, 76, 90]. For example, Montera *et al.* used ANNs to predict wavefront tilt from a Hartmann-Shack WFS [94].

Specialised RNN architectures that simplify the training of ANNs and provide good performance in terms of generalisation and approximation, have been reported in the literature [59]. Recurrent networks include echo state networks (ESNs) and liquid state machines (LSMs). A method will be described in this section that utilises an ESN to predict dynamic phase distortions, representing the SVPSF over a wide FOV. An earlier method using a time-delayed neural network for prediction of aberrations caused by turbulence has been proposed [141]. However, the complexities of training and implementing such networks can be avoided with the use of ESNs. This topic will be outlined in the next section.

5.2 Design considerations

Dynamic modelling of non-linear time-varying systems is achievable using conventional state-space analysis. For example, the state and output equations of a time-varying system can be modelled as

$$\frac{d\boldsymbol{q}(t)}{dt} = \boldsymbol{A}(t)\boldsymbol{q}(t) + \boldsymbol{B}(t)\boldsymbol{x}(t) \tag{5.1}$$

$$\boldsymbol{y}(t) = \boldsymbol{C}(t)\boldsymbol{q}(t) + \boldsymbol{D}(t)\boldsymbol{x}(t), \tag{5.2}$$

where q is the reaction of states at time t, x is the reaction of inputs at time t, and y is the reaction of outputs at time t.

However, the complexity of determining a solution with time-varying coefficients, as shown in Equations 5.1 and 5.2 as $\boldsymbol{A}(t)$, $\boldsymbol{B}(t)$, $\boldsymbol{C}(t)$, and $\boldsymbol{D}(t)$, is compounded by the requirement of *a priori* knowledge of each function with respect to time. In terms of this research, such variations correspond to wavefront phase fluctuations as turbulent air passes over the aperture of a telescope. As a consequence, the spatially variant PSF is affected by the temporal decorrelation of the atmosphere, resulting in image distortion. Adopting Taylor's frozen turbulence hypothesis helps to constrain this property, however it does not fully address the issue of turbulence evolution, as 'frozen' turbulence is moved by winds over the aperture of telescopes that support a wide FOV. ANNs have the capability to address such temporal considerations through training.

Kalman filtering has been used for prediction of wavefront phase perturbations. Such state-space models however, have been limited to AO systems for partial compensation of control lag [105]. As outlined in Section 4.5, this research is based on a DWFS configuration; therefore, to address anisoplanatic imaging, a spatiotemporal image model is required. Employing a Kalman filter [14] to predict spatiotemporal wavefront aberrations was considered, however this was deemed impractical due to complexity of the model required.

Given the requirements of this work, and assuming the Taylor frozen turbulence hypothesis over a constrained FOV, a recurrent neural network was considered appropriate, and is based on:

1. The definition, from a signal processing perspective, of a relatively simplistic configuration that supports a static phase map of Zernike coefficients, moving at a velocity, v, with respect to a fixed, circular, 2D plane (aperture).
2. The ability to train a network for variances, in terms of wind velocity, v_{wind}, and turbulence strength, C_n^2, and to integrate these parameters into a multiple, thin-layer turbulence model.
3. Provision to learn new profiles, commensurate with SCIDAR profiles for specific observation sites, such as MJUO.

It is possible to solve temporal problems using dynamic feedforward neural network architectures [121]. Takens [125] proposed that for dynamic system modelling, the state of

the system can be reconstructed provided that adequate delayed embedding is employed. According to Schrauwen *et al.* [121], such explicit embedding converts a temporal problem into a spatial one. This theory was tested using a dynamic feed-forward neural network architecture using spatiotemporal data.

A simulation platform using a single-layer Gaussian perturbation model was constructed, where 2D spatial time-series data was used to predict x- and y-centroid data over a line segment. A dynamic feed-forward network (DFFN) was trained using spatiotemporal, low-order aberration data from three source beacons. Tilt aberrations were predicted over a line segment representing the path of a science object between two source beacons. These predicted data were compared with actual aberration data acquired over the line segment. In terms of the MSE metric, the results showed a significant reduction in prediction error when the DFFN was employed [141].

However, delay embedding for DFFNs is limited by the length of the delay line used to capture the temporal component of input data. Simple, Gaussian-generated waveforms were used as input for the simulation described in the preceding paragraph [141], however more complex waveforms generated from phase maps would severely limit performance. RNNs, however, support a similar nodal architecture to dynamic feedforward networks, but each internal node supports local feedback, and each network output supports global feedback. Such structures can be used on complex wavefrorms for time-series prediction.

A significant limitation with RNNs is that they lack an efficient training algorithm. Existing algorithms, such as back propagation through time (BPTT) and recurrent back propagation (RBP) are frequently used [85], however since all the weights in a fully (or sparsely) connected network are trained, application of these rules to real-world (complex) problems is not always feasible [121]. In an effort to address such limitations, the so called *random* network was developed; a principal distinction between RNNs and random recurrent networks is that the later remain untrained, only the output weights are processed using a simple regression technique. Further research into such structures resulted in independent inventions of two random recurrent networks, firstly by Jaeger with the *echo state network* (ESN) [59], and secondly by Maass with the *liquid state machine* LSM [82]. The unification of these networks, and similar structures, is referred to in the literature as reservoir computing [121].

The primary difference between ESNs and LSMs is essentially in their application. LSMs support a biologically inspired structure similar to the synapse of the human brain, and use spiking neurons. ESNs are designed as a general engineering tool [81] and act like complex

nonlinear dynamic filters that transform input signals using a high-dimensional, temporal map [121].

For this research, an ESN was developed specifically for predicting phase perturbations represented by Zernike coefficients. Input data comprised angular separations *a priori* and Zernike ensembles *a posteriori* resulting from the output of an extended, curvature WFS; the WFS acquired spatiotemporal image data from several source beacons. A design approach similar to Jaeger was adopted for the ESN, and this is outlined in the following section.

5.3 Echo state networks

5.3.1 Introduction

Echo state networks employ a pseudo-randomly created matrix as a reservoir that remains unchanged during training. ESNs are passively excited by one or more input signals and the state of the network is maintained in the form of a nonlinear transformation of input history [81]. One or more desired output signals are generated as a linear combination of signals generated by sparsely connected nodes within the input-excited reservoir; training, using linear regression, ensures a linear relationship is maintained between the teaching sequence or sequences and the reservoir output.

5.3.2 Architecture

As discussed, ESNs are recurrent networks that simplify training and provide a: "reservoir of rich dynamics" [59]. A fixed sparse matrix, $\mathbf{W}_{\mathrm{DR}}$, is used to implement the recurrent network, and a linear readout, $\mathbf{W}_{\mathrm{out}}$, is trained to produce an output. The state vector, $\mathbf{x}(n)$ maintains the relationship between the input vector, $\mathbf{u}(n)$, and output vector, $\mathbf{y}(n)$, and this is expressed as

$$\begin{aligned} \mathbf{x}(n) &= \varphi(\mathbf{W}_{\mathrm{in}}\,\mathbf{u}(n)^{\mathrm{T}} + \mathbf{W}_{\mathrm{DR}}\,\mathbf{x}(n\text{-}1)^{\mathrm{T}} + \mathbf{W}_{\mathrm{back}}\,\mathbf{y}(n\text{-}1)^{\mathrm{T}}), \\ \mathbf{y}(n) &= \varphi(\mathbf{W}_{\mathrm{out}}\,\mathbf{x}(n)^{\mathrm{T}}), \end{aligned} \tag{5.3}$$

where $\mathbf{W}_{\mathrm{DR}}$ is the dynamic reservoir matrix, $\mathbf{W}_{\mathrm{back}}$ is the feedback matrix, $\varphi(\cdot)$ is a sigmoidal activation function, $\mathbf{W}_{\mathrm{out}}$ and $\mathbf{W}_{\mathrm{in}}$ are input and output weight matrices, respectively, and $n \in \{1, 2, \cdots, N\}$.

The output of the network, $\mathbf{y}(\cdot)$, can be expressed in terms of the input vector, $\mathbf{u}(\cdot)$, as

$$\hat{\mathbf{u}}(n+1) = \mathbf{y}(n), \tag{5.4}$$

where $\mathbf{u}(n+\mathbf{1})$ is a prediction of the input vector one time-step into the future.

With reference to the architecture shown in Figure 5.1, the operation of an ESN can be described as follows. The input vector, $\mathbf{u}(n)$, is mapped into state space, $\mathbf{x}(n)$, through the *echo* property supported by a sparsely connected, recurrent matrix, $\mathbf{W}_{\text{DR}}$. Linear regression is used to train the ESN output matrix, $\mathbf{W}_{\text{out}}$, to facilitate recombination of output data. Input, and optional feedback matrices, $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{back}}$, respectively, are dense, randomly connected matrices, that facilitate the distribution of inputs and output data to the dynamic reservoir. Lastly, as with most ANNs, an activation function $\phi(\cdot)$ is supported to ensure data remains bound and provides non-linear, output capability.

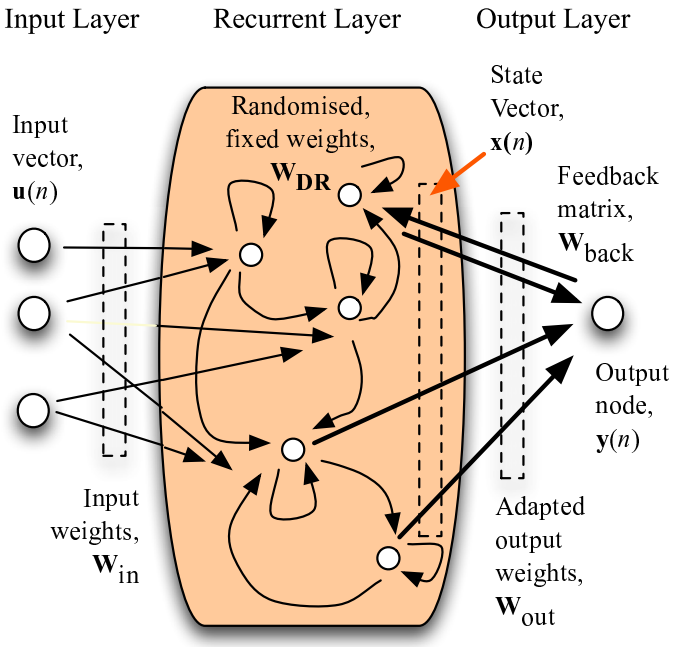

Figure 5.1 Echo state network (ESN) architecture.

For this research an ESN is used as a modal predictor, i.e., given spatiotemporal modal input data from multiple source beacons, modal data, in terms of Zernike coefficients, are estimated for a given target. The extended curvature WFS described in Section 4.4, classifies source image data into modal expansions, i.e., $\{a_2, a_3, \ldots, a_{M+1}\} \leftarrow g(\mathbf{p})$, thus providing

a significant reduction in terms of dimensionality. Using an ESN for both classification and prediction was considered, however this was deemed both inefficient and impractical. Alternative methods to reduce dimensionality for similar problems have been reported [22].

An analysis of results using this ESN architecture for the prediction of Zernike coefficients is presented in Chapter 7. An analysis of ESN design parameters used to optimise network performance is given in next section.

5.3.3 Training

ESN training employs the Moore-Penrose pseudoinverse to adapt $\mathbf{W}_{\text{out}}$ weights for mapping input training vectors, $\mathbf{u}(n)$, to desired output vectors, $\mathbf{y}(n)$. Thus, training requires two data vectors - input data ensembles and expected output data. The input ensembles comprised M Zernike coefficients from K source beacons, defined as

$$\boldsymbol{z}_k = \begin{bmatrix} a_2 \\ a_3 \\ \cdots \\ a_{M+1} \end{bmatrix}, \tag{5.5}$$

where $k \in \{1,2,\cdots,K\}$ source beacons.

Each Zernike coefficient vector (ZCV), $\mathbf{z}_k$, represents the wavefront map from one source beacon over one, discrete-time sample. Each ZCV can be represented in terms of a time-series

$$\mathbf{Z}_k(i\Delta T) = \Big[\mathbf{z}_k(i\Delta T), \mathbf{z}_k(i\Delta T - \tau), \cdots, \mathbf{z}_k(i\Delta T - (N-1)\tau)\Big], \;\; i = n, n-1, \cdots, n-k, \tag{5.6}$$

where ΔT is the sampling period and $k = n-(N-1)\tau$ is the number of reconstructed vectors over a time-series ensemble of length, N.

In addition to the ZCV, an angular separation matrix, $\boldsymbol{\theta}_{\text{Sep}}$, is required to provide spatial reference of training vectors over the pupil. A set of training vectors, $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_R,$, is defined for each training location over the extended pupil. The coordinates of each training vector are combined with the coordinates of each source vector, $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K$, and provide the angular separation between each source and possible target location, such that $\theta_{k,r} = p(\mathbf{s}_k, \mathbf{q}_r)$, where $r \in \{1,2,\cdots,R\}$ and $p(\cdot)$ is the rectangular to polar coordinate operator.

Thus, a separation matrix, $\boldsymbol{\theta}_{\text{sep}}$, can be defined as

$$\boldsymbol{\theta}_{\text{sep}} = \begin{bmatrix} \theta_{1,1} & \theta_{2,1} & \cdots & \theta_{K,1} \\ \theta_{1,2} & \theta_{2,2} & \cdots & \theta_{K,2} \\ \cdots & \cdots & \cdots & \cdots \\ \theta_{1,R} & \theta_{2,R} & \cdots & \theta_{K,R} \end{bmatrix}, \tag{5.7}$$

where K is the total number of source beacons, and R is the total number of training target locations that a science object, T, can occupy over the pupil.

Each row in $\boldsymbol{\theta}_{\text{sep}}$ forms a spatial training matrix (STM), and this was used for ESN training. Both Zernike and separation matrices are combined for ESN training and represent a low-dimensional wavefront map over the pupil. Sequences of wavefront maps over time, form a discrete time-series representing the evolution of turbulence over the pupil. These ensembles were used in a spatiotemporal training algorithm described in the following chapter.

5.3.4 Normalisation and activation functions

During simulation and laboratory testing, the dynamic and static range of Zernike coefficients varied considerably. This was the case for both geometric and curvature WFSs used in data acquisition, and was a function of the Fried parameter, r_0, and intensity of both wavefront images used for acquisition. As a result of considerable range variation, both training and verification data ensembles were pre-processed using normalisation routines, prior to application to the ESN.

To represent an ensemble of Zernike coefficients, the ESN is designed to accept normalised data and uses the hyperbolic tangent transfer function over the range, $[-1,1]$. Most ANN architectures typically use the sigmoid transfer function over the range of $[0,1]$ for the outputs. Normalisation of data vector $\mathbf{u}'$ to produce data vector $\mathbf{u}$, prior to input to the ESN, was performed using the MATLAB® function, `mapminmax` [55].

Sigmoidal activation functions, $\varphi(\cdot)$, are generally used on each processing element within a network to avoid overflow conditions. In addition to sigmoidal functions, linear activation functions can be applied to scale network outputs. A general expression for a commonly

used sigmoidal function, known as the *logistic* function, is [85]

$$\varphi(x) = \frac{A}{1+\exp(-\beta x)}, \tag{5.8}$$

where A is referred to as the nodal gain of the network and is used to either attenuate or amplify the nodal input, β is the gain that determines the overall steepness of the sigmoid function, and x is defined as both vector arguments in Equation 5.3. The constant, A, can influence the overall training time of the neural network [85].

Sigmoidal functions are used as they are continuous and differentiable. Such qualities are required when using learning algorithms such as gradient descent.

For layered, feedforward networks, Kalman and Kwasny [67] recommend an alternative sigmoidal function, the hyperbolic tangent, $\tanh(\beta,x)$, for best overall performance. This function is defined as

$$\varphi(x) = \tanh(\beta,x) = \frac{\exp(\beta x)-\exp(-\beta x)}{\exp(\beta x)+\exp(-\beta x)}, \tag{5.9}$$

Both sigmoidal functions outlined above were used in training and testing the echo state network discussed in Section 5.3.2. The hyperbolic tangent function was found to provide superior performance over the logistics function. As a result, the hyperbolic tangent function was used for $\varphi(\cdot)$ defined in Equation 5.3. An example of a hyperbolic tangent and the logistic sigmoidal functions are shown in Figure 5.2. An ESN can operate virtually as a linear filter if the inputs are close to zero [80]; larger inputs that tend to drive the network into saturation employ the nonlinear region of the hyperbolic tangent function, as shown in Figure 5.2 (a).

In the absence of noise, or for noise levels where the SNR $\geq$ 40 dB, linear predictors (LPs) have been reported to be effective for low-order (tilt) wavefront phase perturbations [76]. However in the same study, Lloyd-Hart and McGuire found ANN predictors using nonlinear activation functions outperformed LPs for low SNR conditions, whilst at least matching the performance of LPs using input data with high SNR [76].

An extensive study has been performed on low-order wavefront phase aberrations using various time-series analysis methods [88]. The results of this study suggest that the wavefront measurement process introduces a low-pass filter condition that enhances predictability of wavefront phase perturbations. Tilt perturbations, however, were only acquired using a H-S

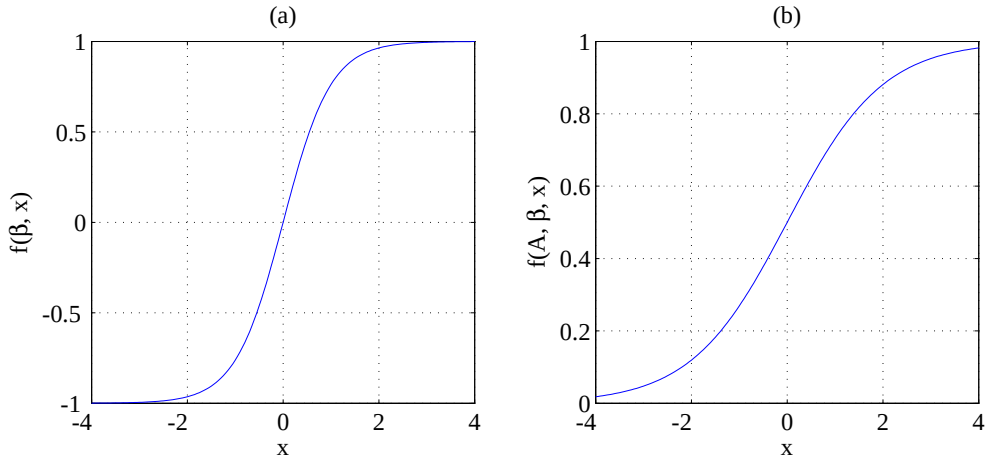

Figure 5.2 Sigmoidal functions: (a) Hyperbolic tangent, $f(\beta,x)$; (b) Logistic function, $f(A,\beta,x)$.

WFS. In a more recent study, temporal analysis was conducted on a range of Zernike coefficients acquired using a curvature-based WFS [142]. The results of this study showed that predictability of Z_4 (defocus), and Z_5 (x-component astigmatism), were improved using ESN and delayed feedforward ANN architectures, both employed non-linear activation functions, compared to linear predictors [142].

Lastly, post network processing was required to renormalise the predicted Zernike data ensembles prior to reconstructing the SVPSF for image restoration. This operation used parameter settings previous stored as part of the normalisation procedure, and was facilitated by the MATLAB® function, `mapminmax`. Thus, the scaling effects on wavefront phase, inherent with the Fried parameter and photon count, were preserved.

5.4 Echo state network optimisations

According to Lukoševičius and Jaeger, a distinction can be been made between traditional approaches to design and application of reservoir computing, and "finer-grained" ideas on producing good reservoirs for ESNs; the former are referred to as 'brands', and the latter, 'recipes' [80]. By this terminology, ESN 'recipes' are explored here.

Given the chaotic behaviour evident in the data ensembles acquired for this study, an artificial neural network architecture was optimised for modal wavefront sequences. This approach is motivated by the generally poor performance of linear predictors when used with chaotic time series [123], and Zernike time series ensembles [142].

The principal use of a trained ESN was to employ spatiotemporal data from source beacons for estimation of phase perturbations within an anisoplanatic patch; such perturbations are indirect measurements of wavefront aberrations from a target object. Simulated wavefront data is used to train the ESN. To optimise ESN efficacy, system performance is measured with actual wavefront sequences. This is the principal topic addressed in this section.

5.4.1 Design parameters

The performance of an ESN can vary considerably and typically require tuning using global controls to meet design objectives. Such controls are termed *design parameters* and are used to specify architectural constraints, for example, network size and dynamic sensitivity. A summary of parameters used for the design of the ESN are:

- The input dimension, U.
- Number of neurons in the dynamic reservoir.
- The density of connections within the dynamic reservoir.
- The spectral radius, $\rho(\mathbf{W}_{\mathrm{DR}})$.

These parameters are defined by Jaeger *et al.* [60], and a brief summary of each is given here. The input dimension, U, is simply the number of inputs presented to the network. In this study, each input represents either an angular displacement or a Zernike coefficient; combined, these represent spatiotemporal wavefront perturbations of source beacons. The input dimension, U, is defined by Equation 6.25 on Page 141, and increases linearly as more source beacons are added.

The number of neurons is a measure of the reservoir size. The dynamic reservoir (DR) is implemented as a 2D square matrix, however only one dimension, i.e., column or row size, is typically used to define this parameter. The density of connections within the dynamic reservoir refer to recurrent pathways within the DR [57]; such pathways are used to form a recurrent, signal connection. Related to this, reference is also made to the sparsity or *sparseness* of the network. Lastly, the spectral radius, $\rho(\mathbf{W}_{\mathrm{DR}})$, is the magnitude of the largest eigenvalue within a sparse matrix.

For the majority of ESN applications, and in this study, values for these parameters are statically defined, i.e., parameters remain constant during network operation. Recently,

an alternative approach has been reported, where a genetic algorithm dynamically alters network parameters in response to variations in input dynamics [149].

An ESN architecture described in Subsection 5.3.2 was used for this analysis. However, because the calculation of the MSE is computationally expensive when modifying multiple parameters, network size was reduced. The size of the dynamic reservoir was reduced commensurate with the test input dimension. The smaller implementation allowed a wider range of tests to be conducted over the evaluation period, whilst maintaining the same overall structure of the ESN defined in Chapter 7. Effectively, the ESN used in this analysis was a scaled version of the ESN predictor used in this study.

Early tests suggested that a large, densely connected network adversely affects performance. This was confirmed using two inputs, Zernike coefficients, a_2 and a_3, where only temporal performance, as opposed to spatiotemporal prediction, was considered. Spectral radius was fixed, where $\rho(\mathbf{W}_{\mathrm{DR}}) = 0.4$, and 50 ESNs were generated, trained, and tested, ranging in size from 6 to 35 neurons. The results for single time-step prediction using 300 time-samples from simulated training dataset, `Sr0_20z10kT1`, and verification dataset, `Sr0_20z10kV1`, is shown Figure 5.3.

For each given DR size, 50 ESNs were generated, trained and simulated. The mean was calculated for each reservoir size and is shown as a continuous line in Figure 5.3; both the minimum and maximum values are also shown as error bars. These results suggest the optimal network size required to achieve an average MSE of approximately 0.065, is 24. This reservoir size also returned the lowest error variance, as shown by the error bar between 0.062 and 0.068. Given the input requirements discussed earlier in this section, the dynamic reservoir supported approximately 600 neurons for spatiotemporal prediction.

Additional parameters were varied and assessed using tilt time-series ensembles from the dataset, `Sr0_20z10kT1`. For example, the spectral radius, $\rho(\cdot)$, was also varied 0.1 and 0.8. The averaged results of this performance analysis, in terms of normalised MSE, are shown in Figure 5.4.

Jaeger suggested that for optimal performance, the spectral radius should operate on the edge of stability, i.e., approaching, but not equal to unity [59]. To verify this, the spectral radius, $\rho(\mathbf{W}_{\mathrm{DR}})$, was iterated at 0.1 intervals, such that, $0.1 \leq \rho(\mathbf{W}_{\mathrm{DR}}) < 1.2$, and a statistical MSE average was calculated over 50 iterations. For values $0.4 \leq \rho(\mathbf{W}_{\mathrm{DR}}) < 0.8$ the network performance was reasonably consistent. However, when $0.98 < \rho(\mathbf{W}_{\mathrm{DR}}) \leq 1.2$,

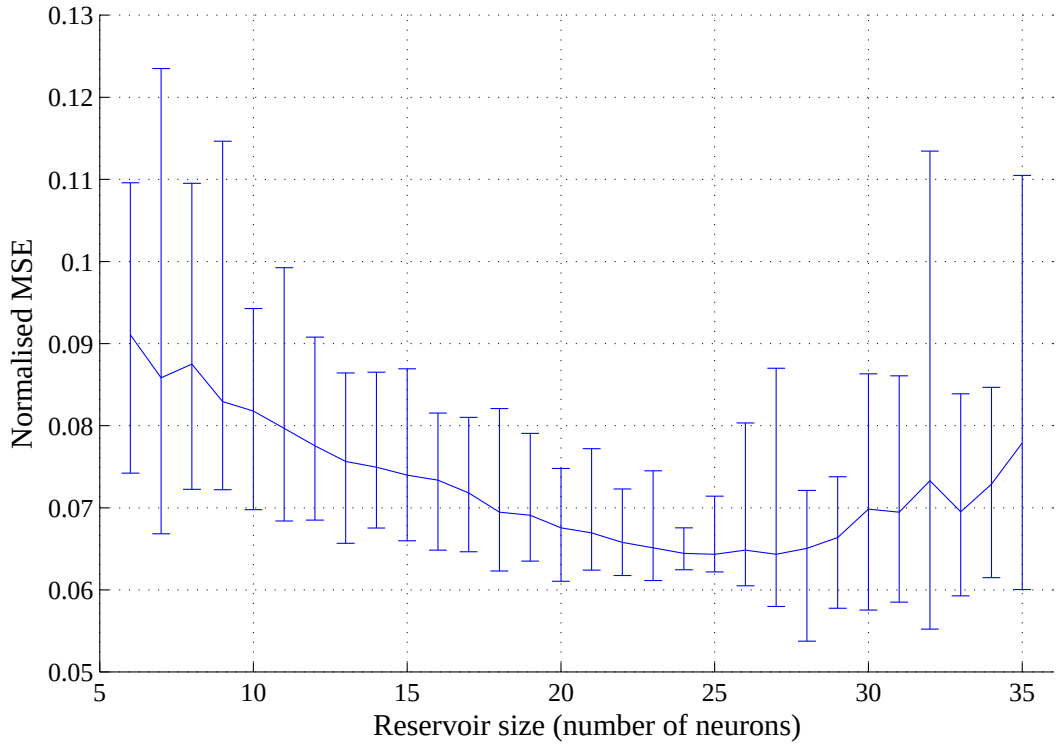

Figure 5.3 Averaged MSE of tilt predictions over a range of network sizes.

the network became increasingly unstable.

Network performance was also assessed as a function of neurons and $\rho(\mathbf{W}_{\mathrm{DR}})$. The normalised MSE in Figure 5.4 increases as the size of the dynamic reservoir exceeds 25 neurons; this increase is significant when combined with spectral radius values, $0.3 > \rho(\mathbf{W}_{\mathrm{DR}}) \geq 0$.

Lastly, $\rho(\mathbf{W}_{\mathrm{DR}})$ was varied with network sparseness, and the MSE evaluated. Figure 5.5 shows an overall performance gain between spectral radius values $0.1 \leq \rho(\mathbf{W}_{\mathrm{DR}}) < 0.9$, converging to MSE ≈ 0.07, as ESN density is reduced.

Given these results, the following items were considered necessary to achieve optimal performance gains:

1. A sparsely connected network allows for relative decoupling of subnetworks, and encourages development of internal dynamics [57]. Given the complexity of the time

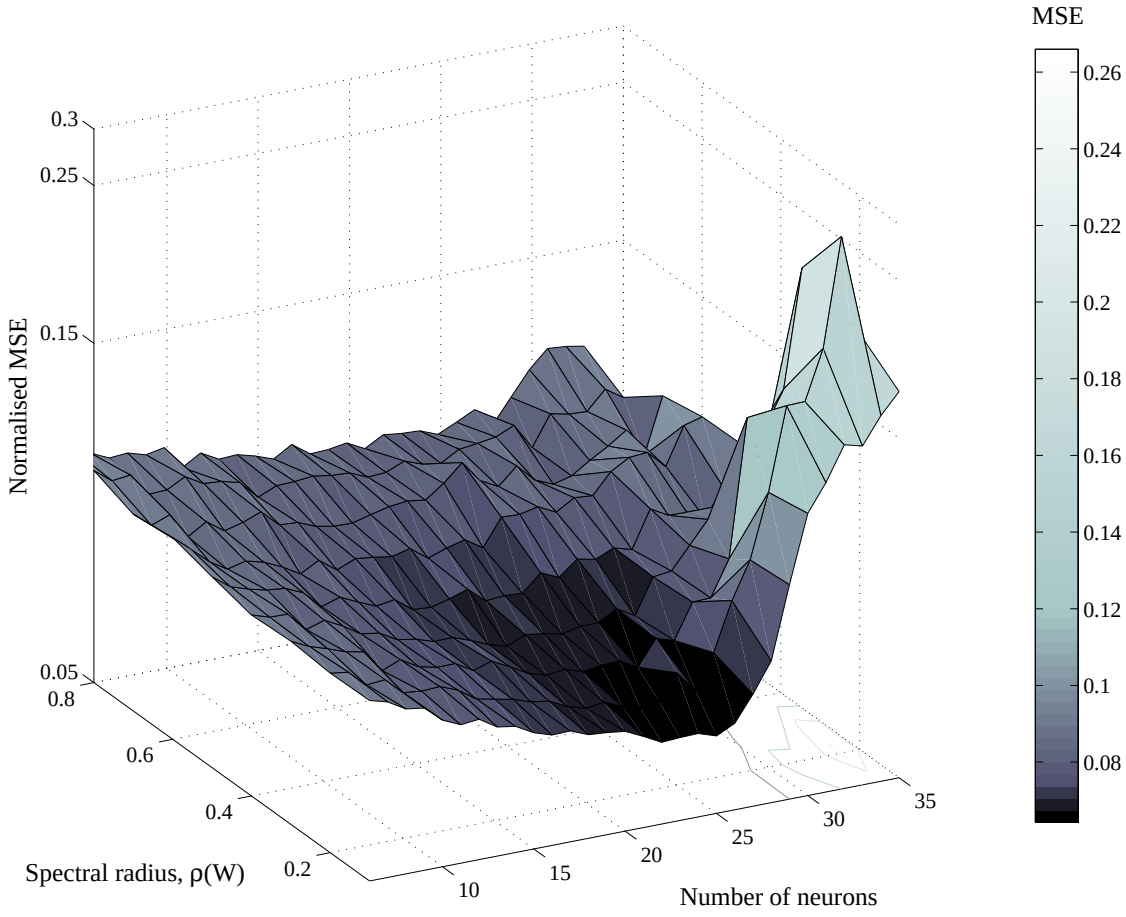

Figure 5.4 ESN performance by spectral radius and the number of neurons.

series input sequences, a rich dynamic structure is required. Sparsity values of 0.1 appear to meet this requirement, as shown in Figure 5.5.

2. Results from the time series analysis presented in Table 4.2 confirmed high embedding dimensions, D_E, for Zernike coefficients. Such a property requires a long memory for prediction. To address this requirement, the spectral radius, $\rho(\mathbf{W}_{\mathrm{DR}})$, should be as high as possible, but without violating the echo state property [57] for reservoirs that use the tanh function, and for zero input [80]. In this case, we chose $\rho(\mathbf{W}_{\mathrm{DR}}) = 0.95$.

3. A sufficient, however not an excessive, number of neurons are required to support the recurrent pathways within the dynamic reservoir. Defining an excessive number of neurons can result in severe performance degradation when combined with low

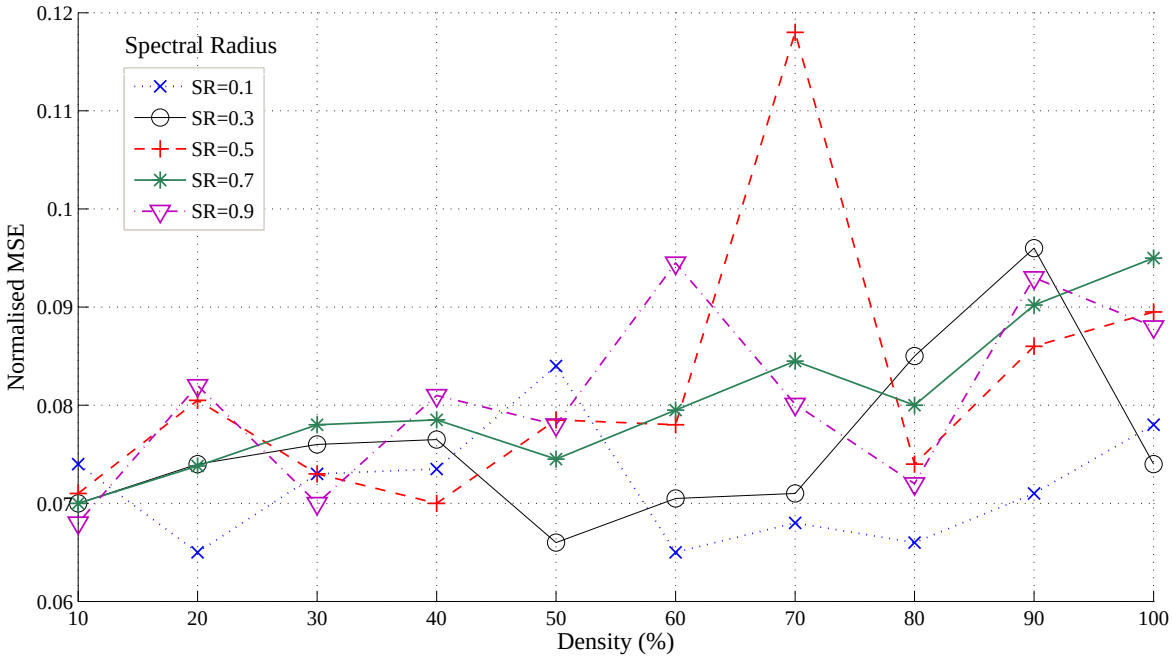

Figure 5.5 ESN performance by spectral radius and sparseness.

$\rho(\mathbf{W}_{\mathrm{DR}})$.

4. Ensuring a proportional quantity of neurons, commensurate with input dimension, U, is required to prevent saturation of the dynamic reservoir (this consideration is expanded in Subsection 7.7.1, where different combinations of slightly correlated inputs are analysed).

These results are consistent with recommended ESN configurations reported in the literature [80,102]. Thus, design parameters were assigned values for implementation of an ESN used in this study. A summary of these design parameters is given in Table 6.2 on Page 143.

5.4.2 Performance evaluation

The performance of an optimised ESN was evaluated for single time-step prediction of the simulated dataset, `Sr0_20z10kV5`. ESN parameters were selected, based on results from the analysis described in this section, and comprised a single input, U, a dynamic reservoir of 22 neurons, a connection density of 0.1, and spectral radius, $\rho(\mathbf{W}_{\mathrm{DR}}) = 0.95$. The ESN was trained using datasets acquired during the successive simulation runs, i.e.,

under similar conditions. Single time step prediction of tilt perturbations, $\{a_2, a_3\}$, were individually assessed using a trained ESN. Formatted training sequences, comprising time series data ensembles, were acquired from successive runs.

A data sequence of 400 samples was evaluated, with a washout time[1] of 100 samples. ESN training was performed using 3,000 data samples. The training methodology was based on supervised learning, and verification datasets were used to assess performance. Based on this configuration, a minimum mean squared error (MSE) of 0.05421 was measured for single time-step prediction. The normalised results from each test run, employing ESNs trained and optimised on individual Zernike coefficients, are shown in Figure 5.6.

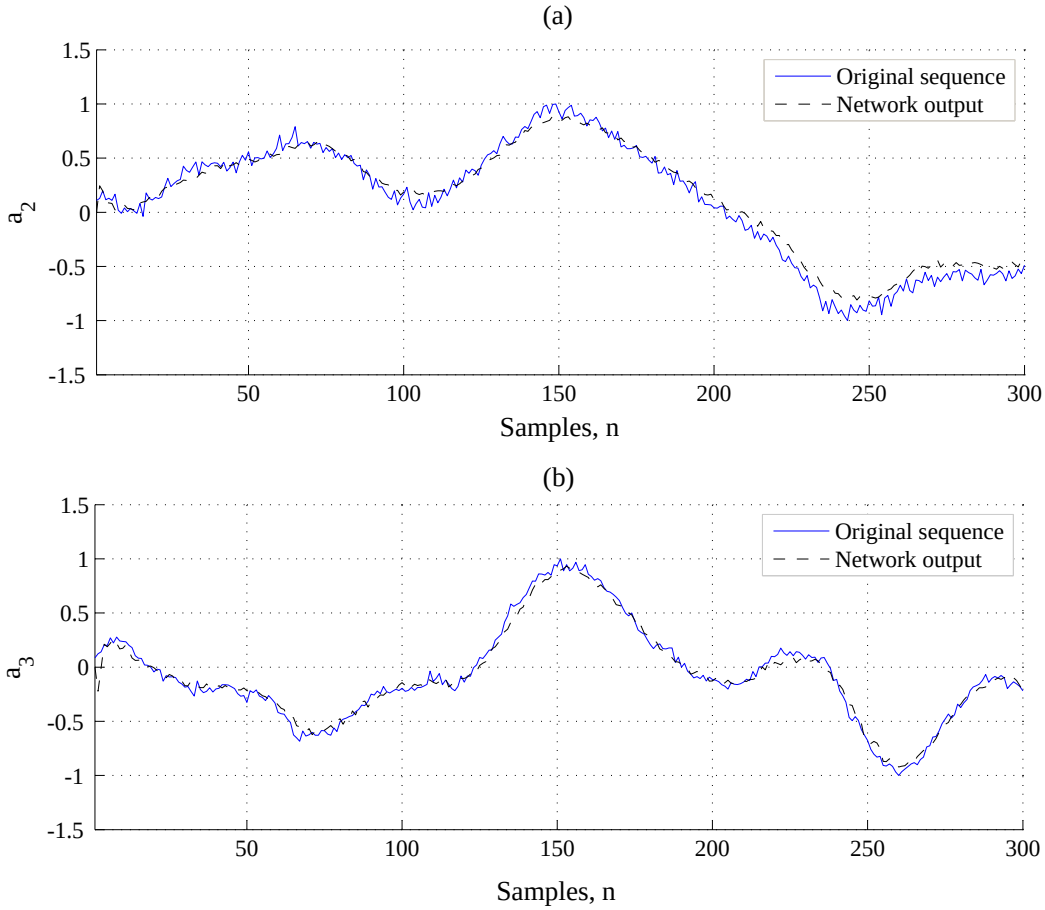

Figure 5.6 Single time step prediction of tilt perturbations, each output from an ESN configured using optimal design parameter values; (a) Zernike coefficient a_2; (b) Zernike coefficient a_3.

[1] *washout time* is used to define a period, typically in terms of time-samples, for the network to stabilise, prior to commencement of performance assessment.

Inspection of subfigures (a) and (b) in Figure 5.6 shows the network prediction of each time sample closely tracking the original series. Over the 300 time sample segment, this resulted in MSEs of 0.09392 and 0.05421, respectively. Of particular interest are maximums and minimums of the original time series, where higher errors are encountered. Examples are shown in Subfigure (a) at sample points 100, 150 and 250, and to a lesser extent in Subfigure (b) at sample points 70, 260.

5.5 Noise and memory considerations

A principal source of noise in astronomical images is due to photon starvation, however if CCDs are employed, either for direct imaging or for wavefront detection, additive read noise must also be considered. For this work, CCDs were used for DWFS; this required both photon and CCD read noise for simulations, generated by Gaussian and Poisson noise distributions, respectively. To simulate wavefront propagation and measure phase perturbations, wavefront images were distorted using a range of noise levels (SNR in the case of read noise and a photon count for low-light conditions). However, different compositions of noise from various sources affected the performance of the ESN in various ways.

For predicting low-order wavefront phase perturbations, limited additive noise during training has been shown to improve network performance when verified using noisy test data [76, 90]. However, only Gaussian noise was used in this study - the effects of Poisson noise on the prediction of low-order aberrations using ANNs remains unreported. General studies conducted on ESNs using additive noise have reported encouraging results, i.e., network stability is increased, especially for configurations that support recurrent outputs [58, 60].

A fully recurrent configuration, where network outputs are feed back as network inputs, was not used in this study. When tested, global feedback resulted in degradation of overall network performance; this applied to each Zernike mode tested. Conjecture is based on the structure of the Zernike time-series sequences; shorter signal responses were more easily learnt by the network, compared to longer sequences that would require global feedback. This suggests that network memory capacity without global feedback was adequate to learn the perturbation pattern for each mode, however enabling global feedback may introduce instability, as has been reported in the literature [60].

After initial testing, where wavefront intensity images were distorted with additive Gaussian noise, $40 \leq \text{SNR} \leq 60$ dB, and the ESN was trained using those random data en-

sembles, the predicted MSE wavefront error showed marginal improvement. However, as somewhat expected, this beneficial result was not repeated in the presence of photon noise. The results of these performance conditions are presented in Section 7.5 and 7.6, respectively. In terms of this network architecture however, these improvements are thought to be related to the unintentional introduction of regularisation, where a small proportion of additive noise results in ridge regression; this is also known as Tikhonov regularisation [9]. Such conditions will prevent the network from over fitting the exact examples given in training, resulting in better generalisation of the network.

The effects on the memory capacity of the network was also considered. Initial results on several, low SNR studies, showed a significant increase in AMSE; a relationship between the noise conditions and short-term memory was considered. For example, a reduction in SNR may result in signal memory being washed out by noise. This will lead to a shorter effective memory of the signal. Since optimal performance for ESNs is still based on a manual refinement of the system, this topic, and the general shaping of the network by fine-tuning several global control parameters, is discussed in the preceding section.

5.6 Prediction vs. estimation

Given the most recent sample of a discrete, time varying signal, $\zeta(n\Delta t)$, and a record of past events $\zeta(n-\Delta t), \zeta(n-2\Delta t), \cdots, \zeta(n-N\Delta t)$, prediction can be defined as stating the outcome of an event, $\hat{\zeta}(n+T\Delta t)$, T sample periods into the future. For example, the function performed by the ESN given by Equation 5.4, where the output vector $\mathbf{y}(n)$ is defined in terms of a current, and a series of historical events, $\mathbf{u}(n), \mathbf{u}(n-1), \cdots, \mathbf{u}(n-N)$, prediction, one time-step into the future, is defined by $\hat{\mathbf{u}}(n+1)$.

In terms of the generalised image model developed for this study, both temporal and spatial data are combined using a trained ESN to estimate the spatially varying point spread function (SVPSF) for anisoplanatic image restoration. As a result, since angular separations and temporal data are combined to define $\hat{\mathbf{u}}(n)$, *prediction* of the next outcome of these data ensembles is an appropriate term. However, from the perspective of reconstructing the SVPSF using predicted ESN data ensembles, and given noise sources typically used for astronomical imaging, *estimation* of the degradation function over the image plane is the appropriate term. In this dissertation, both terms will be used where appropriate.

Chapter 6

Simulations

A platform for the study of anisoplanatic effects on the space varying point spread function (SVPSF) is described in this chapter. A simulation framework was developed that comprised two components, a wavefront propagation model, and a target predictor. The propagation model generated data ensembles that were statistically equivalent to spatiotemporal wavefront phase, perturbed by turbulence. Each ensemble represented a source beacon acquired by measuring wavefront curvature through a 1-m class telescope. The temporal evolution of turbulence over the telescope aperture was mapped using time series data, simultaneously acquired from multiple source beacons.

The second component of the framework comprised a predictor, based on an echo state network (ESN). Data ensembles generated from the propagation model were used to firstly train the ESN to predict perturbed wavefronts of a target over an extended region, where field angles between target and source beacons exceeded the isoplanatic angle. Once trained, the ESN architecture was used to predict perturbed wavefronts from targets over anisoplanatic regions. Phase perturbations were represented by the Zernike modal expansion, where the prediction of individual wavefront aberrations could be assessed.

Monte Carlo simulation was used to conduct pseudorandom explorations of solution space. Monte Carlo methods have been used to solve many complex problems, such as inverse imaging [1] and in modelling the atmospheric turbulence problem using stochastic processes [3]. However, to find such solutions, considerable computer resources are required. To address this requirement, the Beowulf cluster [99] was employed for the simulations outlined in this chapter.

A general and flexible modelling environment was found in MATLAB® computing software [55]: existing phasescreen generation [63] and curvature sensing routines [19] were available. These routines were reviewed and where appropriate, adopted for this study. In addition, several MATLAB® toolboxes were used to build the simulation models described in this chapter; these included the statistics, neural network, and image processing toolboxes [55]. Integration of a specialised AO MATLAB® toolbox was investigated [101], however due to licensing restrictions this package was not available for use outside the United States of America.

This chapter is organised as follows. An overview of the simulation framework is presented in the next section. The two primary simulation components that comprise the framework are discussed in detail. Section 6.2 describes the propagation simulation environment. The predictive simulation environment is described in Section 6.3, where formulation of the data structures for application of the spatiotemporal training algorithm is given. A verification model is also presented in Section 6.3; this model is employed in Chapter 7 to evaluate the ESN method.

6.1 System model

Since independent parameters were used in each framework, simulations were conducted in three stages. The first stage required the propagation of broadband light, in the visual and near infrared spectrum, from source and target beacons using the phasescreen method developed by Harding *et al.* [47]. The geometric wavefront sensor originally developed by van Dam and Lane [136], and analysed by Yong [19, 20] was extended and used to extract a series of Zernike coefficients, representing phase perturbations from imaging multiple source beacons through turbulence. As one or more phasescreens were moved over a fixed aperture, time-series data ensembles were generated for processing by a predictor network. Those data ensembles were employed in the second simulation stage for training an ESN for prediction of phase perturbations over wide, angular separations between target and multiple source beacons. Verification was performed using actual target data, and the AMSE metric was used to assess the results presented in Chapter 7. Predicted, modal perturbations, in the form of Zernike coefficients, were then used to represent the SVPSF. The final phase of simulations required application of the estimated SVPSF for anisoplanatic image restoration of target objects over a wide field-of-view. This topic is addressed in Appendix A.

A system model, based on the research conducted in this thesis, is shown in Figure 6.1. The image of a science object, $\mathbf{f}_T$, is perturbed by atmospheric turbulence, and a distorted facsimile of the original object is projected on to the image plane as $\mathbf{g}_T$. A telescope and imaging device, such as a CCD camera, are used to view and record the effects of atmospheric turbulence on three source beacons and a target object over a wide field-of-view (FOV). These turbulent effects can be measured in the pupil plane as inhomogeneous phase distortions. The system requirement is to use the images of three aberrated source objects to restore the image of the faint target object, $\mathbf{g}_T$, over field angles in excess of 100 μrad.

To achieve this, the pupil is divided into zones, each zone corresponds to a region-of-interest (ROI) at the image plane. Three ROIs, representing the effects of turbulence on three source beacons are designated, $\mathbf{g}_1$, $\mathbf{g}_2$, and $\mathbf{g}_3$. Phase perturbations, corresponding to the effects of distortions from each source beacon are simultaneously captured and translated into modal expansions using a WFS, i.e., Zernike coefficients, $\mathbf{z}_1, \mathbf{z}_2$ and $\mathbf{z}_2$, where each vector is defined by Equation 5.8.

A trained echo state network (ESN), in conjunction with angular separation data, $\boldsymbol{\theta}_{\text{sep}}$, is used to predict phase aberrations, $\mathbf{z}_{\text{T}}$, perturbing an ROI representing a target object $\mathbf{f}_{\text{T}}$, relative to angular separations provided by $\boldsymbol{\theta}_{\text{sep}}$. From these data ensembles, the distortion operator, $\mathbf{H}_{\text{T}}$, can be recovered. Lastly, a deconvolution algorithm is used in conjunction with the distorted image of the target object, $\mathbf{g}_{\text{T}}$, and $\mathbf{H}_{\text{T}}$, for restoration of the target image, $\hat{\mathbf{f}}_{\text{T}}$.

The dimensionality of the problem is significantly reduced by the extended geometric wavefront sensor. Each Zernike mode represents the eigenvalues of phase aberrations perturbing the image from each source beacon. Thus, the ESN, shown as the *modal predictive network* in Figure 6.1, is trained using perturbations obtained from several thousand training vectors, collectively representing various turbulence profiles over an extended pupil. The ESN will respond with target modal data if the network has been trained with a similar pattern to that which has been presented at the input. The evolution of the perturbations over the FOV is mapped using time-series ensembles of modal aberrations. Given similar, temporal patterns that have been trained into the ESN, the output of the system is essentially driven or *steered* using the angular separation data, to predict a target pattern,given $\boldsymbol{\theta}_{\text{sep}}$, as defined by Equation 5.7. The network output is composed of modal aberrations, corresponding to the spatial region specified by each angle defined in $\boldsymbol{\theta}_{\text{sep}}$. Given noise, such as additive read noise and photon noise, the ability of the network to perform using field observation data is of particular interest. Considering these adverse conditions, the performance of the

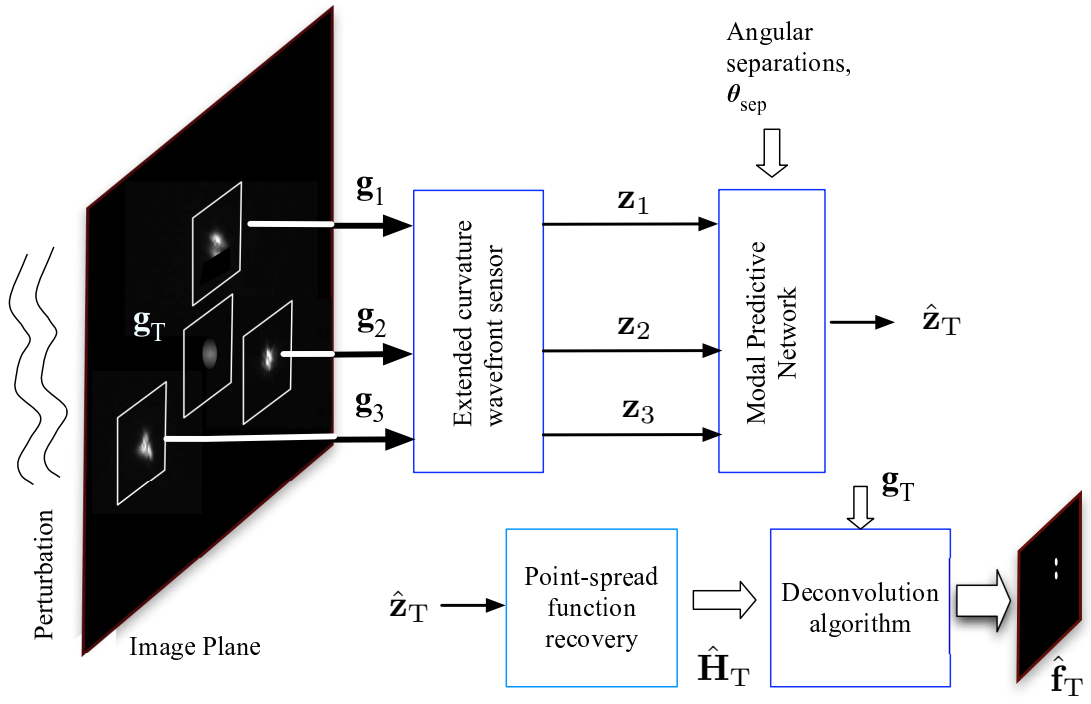

Figure 6.1 System model for predicting the spatially variable PSF.

predictor is assessed in Chapter 7.

6.2 The propagation model

Atmospheric turbulence is typically viewed as a thin-layered structure, generally confined to the troposphere and lower stratosphere, i.e., 0-12 km in altitude; above the tropopause, the effects of turbulence rapidly decrease and disappear above 25 km in altitude [48]. This view is incorporated in the propagation model developed as part of this study, and is discussed in this section. A framework is provided for simulating wavefront propagation from multiple source and target objects, where the effects of phase perturbation through several turbulent layers are studied. Representation of these effects in the image plane, and acquisition of the resulting aberrations using curvature-based wavefront sensors, is included in the model.

6.2.1 Overview

A model described by Roggemann and Welsh [117] for deconvolution from wavefront sensing (DFWS) was reviewed for use in this research. However, this was limited in several respects. Firstly, simulations provided only one reference beacon, essentially to model the effects of classical AO. Secondly, the model was designed to acquire spectral statistical data; however, the ability to simulate using temporal parameters to assess the temporal dynamics of the simulation, is a major requirement of this research. Lastly, and more generally, a single, stationary phase screen was generated to image one source and target object; more versatility was required to extend the simulation environment, suitable for anisoplanatic imaging. An adaptive optics modelling package was also reviewed for possible use in this study [16], however the WFS definitions were restricted to either the H-S or Pyramid configurations, thus limiting the number of sources to one beacon.

Simulating wavefront propagation over multiple layers of turbulence is often a complex and computationally expensive procedure. As a result, three simplifications were applied to this task. Firstly, the optical effects of turbulence can be calculated by summing along straight line ray paths, as is done in geometrical optics. This approach ignores the effects of diffraction and simplifies wavefront propagation used in simulations [61]. Such a condition can be assumed for relatively weak turbulence found at good astronomical sites, or for relatively short propagation paths [49]. The second simplification is to assume that the refractive index of air remains constant between layers, i.e., a wavefront can be considered to pass through a vacuum. Lastly, diffraction effects can be neglected in a thin-layer model, since there is insufficient distance for such effects to evolve within the layer [63].

These simplifications were incorporated in the model shown in Figure 6.2. The geometry comprises a two-layer propagation model, where α, β, and γ are natural guide star (NGS) source beacons, and ζ is a target (science) object at zenith, with angular separations between the target and each source beacon of, $\theta_{\alpha\zeta}$, $\theta_{\beta\zeta}$, and $\theta_{\gamma\zeta}$, respectively. An observational platform G, and two turbulent layers, L_2 and L_1 at altitudes h_2 and h_1 respectively, are shown, with corresponding refractive index structure constants, $C_n^2(h_2)$ and $C_n^2(h_1)$.

Given the propagation model shown in Figure 6.2, a phase estimate of the photon-starved target object, ζ, is calculated using one or more bright NGSs, α, β, or γ. Consider now the isoplanatic angle, θ_0, such that, $\theta_{\gamma\zeta} \leq \theta_0$, $\theta_{\gamma\alpha} > \theta_0$, and $\theta_{\gamma\beta} > \theta_0$. Since wavefronts $W_\zeta(x,y)$ and $W_\gamma(x,y)$ pass through the same isoplanatic patch over both layers L_2 and L_1, phase estimates measured from γ are considered effectively equivalent to ζ. However,

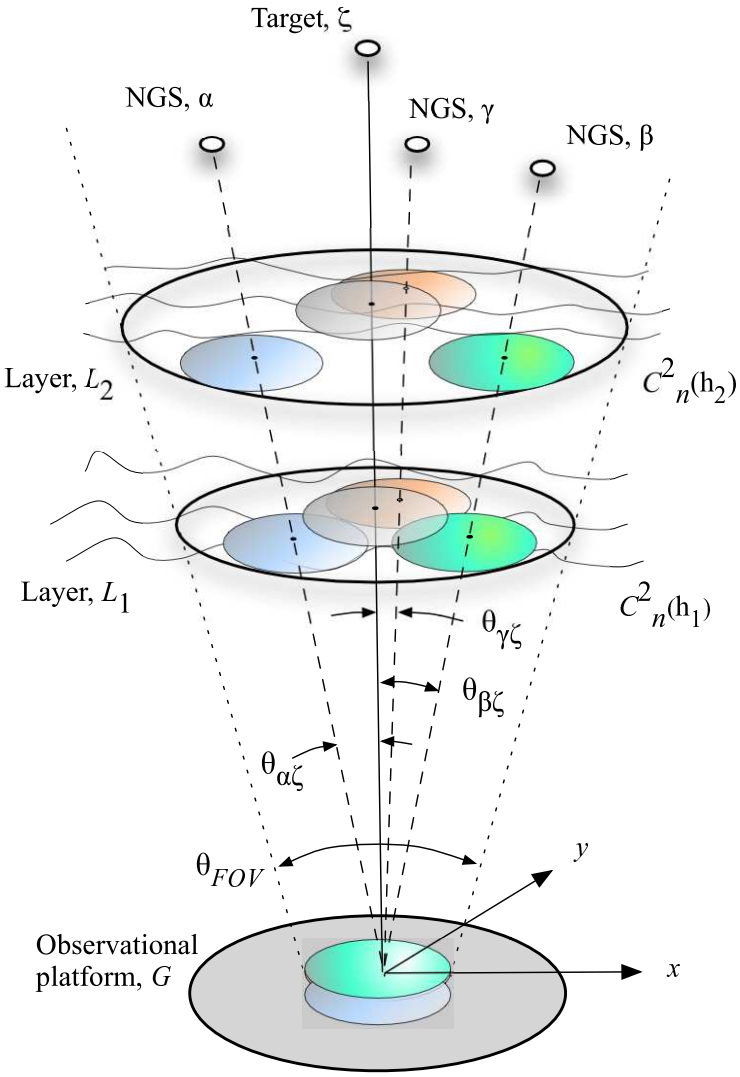

Figure 6.2 Geometrical view of the simulation propagation model showing the isoplanatic region, where $\theta_{\gamma\zeta} \leq \theta_0$.

phase measurements acquired from either α or β would result in a poor estimation of phase perturbations distorting target object, ζ.

With reference to Figure 6.2, the anisoplanatic imaging problem can be restated, such that, simultaneous phase measurements from multiple anisoplanatic source beacons, e.g., $\phi_\alpha(\mathbf{x}_1)$ and $\phi_\beta(\mathbf{x}_1)$, can be used to estimate wavefront phase, $\hat{\phi}_\zeta(\mathbf{x}_1)$, over a wide FOV for image restoration of a distorted target, $\hat{\mathbf{f}}_\zeta(\mathbf{x}_2)$. Various methods, such as outlined for DWFS and CDWFS systems, are effective in reducing the effects of anisoplanatic imaging however, a new method is proposed that uses a trained ESN for prediction of the SVPSF; this method is presented in Section 6.3.

6.2.2 Propagation model parameters

The isoplanatic angle, θ_0, is considered one of the most important parameters in the propagation model; it is typically between 5-10 μradians in the visible spectrum [90]. Angular separation, θ, that exists between a target object and one or more source beacons, is defined over the range, $0 \rightarrow 242\,\mu$rad. Thus, the usable FOV, given a target object orientated within the centre of the pupil, is $0 \leq \theta_{\text{FOV}} \leq 484$ rad.

Parameters used to define θ_0 were used as model variables. For example, originally defined in Equation 4.16 θ_0 can also be defined in terms of Fried's coherence length, r_0, [116]

$$\theta_0 = 0.314\frac{r_0}{\bar{h}}, \tag{6.1}$$

where $\bar{h}$ is the average height of the turbulent layer, and can be defined in terms of the turbulence structure constant, C_N^2,

$$\bar{h} = \left[\frac{\int_0^L C_N^2(z)z^{5/3}dz}{\int_N^2 C_n^2(z)dz}\right]. \tag{6.2}$$

The Fried parameter ranged between 0.05 m for relatively strong turbulence, to 0.25 m for relatively weak turbulence. In addition, the height of the turbulence, h, ranged between 0 m and 12×10^3 m. However, when using the ESN method, empirical data measurements imposed a minimum practical limit of 2 km.

Variations in wavelength, λ, for use in either laboratory or field application, were also considered in the propagation model. For example, laboratory equipment used laser sources, where $\lambda = 625 - 640$ nm, however CCDs were optimised for $\lambda = 400 - 700$ nm for CCDs, where QE $= 0.2$[1], thus, the simulation propagation model was configured for wavelengths of $400 - 680$ nm.

Lastly, since a spatiotemporal model is used, phase screens were moved with independent velocities over a stationary aperture. As a result, both the exposure period and sampling rate (equivalent of frames-per-second) are also qualified. The minimum and maximum values of parameters used for propagation simulations are listed in Table 6.1.

[1] Kodak CCD image sensor, KAI-0340DM

Table 6.1 Propagation model parameters.

Propagation parameter	Values	
	Min.	*Max.*
Fried Seeing Parameter, r_0, m	0.05	0.25
Focal Ratio, f	13.5	13.5
Aperture diameter, D, m	1	1
Optical peak wavelengths, λ, 10^{-9} m	500	650
Usable optical bandwidth, λ, 10^{-9} m	600	900
Wind velocity, $\mathrm{m\,s^{-1}}$	2	25
Propagation distance through atmosphere, 10^3 m	2	12
Number of phase screen iterations	100	1000
Angular separation, θ, 10^{-6}, rad	0	242
Sample spacing in the pupil plane, 10^{-2} m	0.65	0.65

6.2.3 Phase screen generation

A turbulent layer of air can be described in terms of a *random phase screen* if the phase of the transmitted light is changed in an unpredictable fashion and does not absorb light [43]. The amplitude transmittance can be defined as

$$U_{\text{trans}}(x,y) = A\,\exp[j\phi(x,y)], \tag{6.3}$$

where $\phi(x,y)$ is the random phase shift and A is the amplitude of the incident monochromatic plane wave.

The amplitude transmittance is also the phase component in the generalised pupil, $\mathscr{P}$, given by Equation 3.5. To detect phase perturbations in the generalised pupil, a method was required to generate $\phi(\cdot)$ for simulations.

The random midpoint displacement method [47] was used to generate phase screens used for simulation. Each phase screen was based on a telescope diameter, D, of 1 m, and Fried

coherence length, r_0, of between 0.05 m and 0.25 m. This resulted in a range of turbulent profiles, where $D/r_0 = \{4, \cdots, 20\}$. Each generated phasescreen provided Kolmogorov power spectrum of refractive index fluctuations over a wide FOV as described. An example of a generated phase screen, with $D/r_0 = 10$, is shown in Subfigure 6.5 (a).

Since the Kolmogorov turbulence model is used in the creation of the phase screen [62], the piston term was removed. The piston term has an infinite covariance DC component representing phase fluctuations in object space [62]. Wavefront phase, shifted by propagation through a phase screen over a finite area defined in terms of a region from the projected pupil, is represented by removing the mean, such that, [62]

$$\phi(x,y) = \psi(x,y) - \int\int W_A(x',y')\,\psi(x',y')\,\mathrm{d}x'\,\mathrm{d}y', \tag{6.4}$$

where $\psi(\cdot)$ represents wavefront phase prior to removal of piston, $W_A(\cdot)$ is the weighting function defining the physical extent of the phase screen [117],

$$W_A(x',y') \begin{cases} c & \text{over the region of simulation} \\ 0 & \text{outside the region of simulation,} \end{cases} \tag{6.5}$$

and where the constant, c, is chosen such that

$$\int\int W_A(x',y')\mathrm{d}x'\,\mathrm{d}y' = 1. \tag{6.6}$$

6.2.4 Plane wave propagation through turbulence

As described in Section 3.1, a wavefront, $W(x,y)$, can be described in terms of optical path length (OPL) using geometrical optics. However, geometrical optics breaks down when light rays intersect. Diffraction is used to describe an intensity distribution, such as a focused image of a NGS. In this subsection, the simulation of a planar wavefront, incident on a perturbing medium and propagated to the focal plane for image formulation and subsequent wavefront estimation, is described. The associated mathematical processes are defined for both single- and multi-layer turbulence models.

Single layer propagation

A phase screen is used to simplify wavefront propagation, where a thin, random screen can effectively replace a turbulent region [117]. Kolmogorov statistics are generated using the

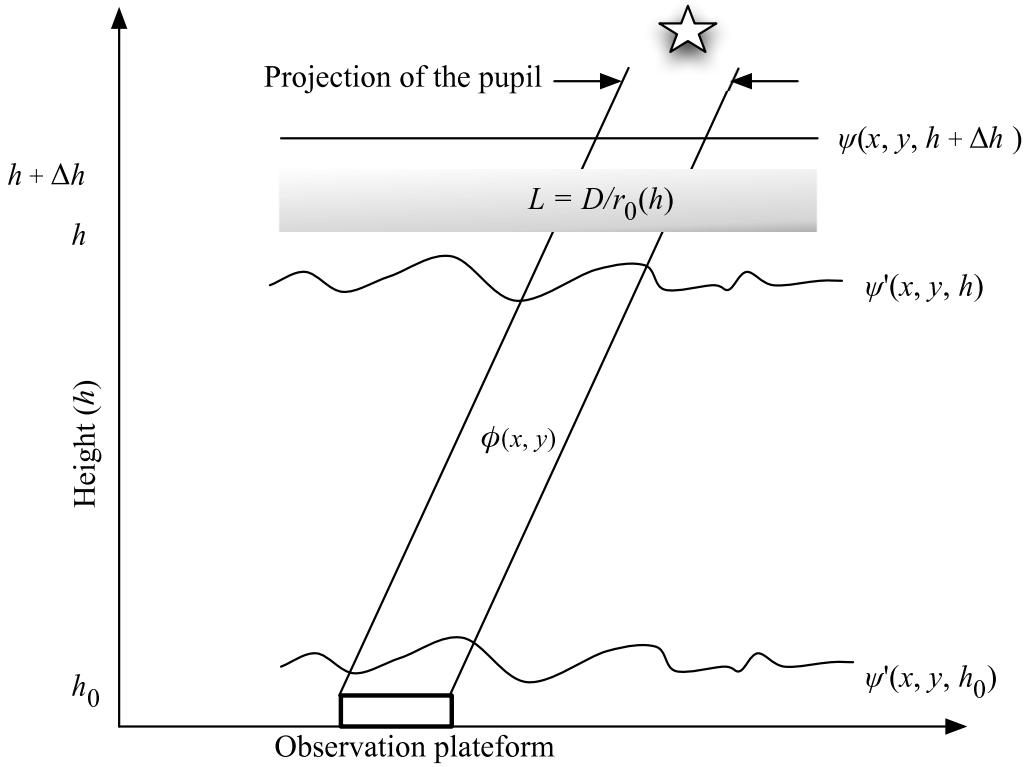

Figure 6.3 Plane wave propagation through a single turbulent layer.

mid-point algorithm [47] and the size and strength of the phase screen are characterised by Fried's parameter, r_0, the diameter of the aperture, D, and height of the turbulence, h. Each screen model represents the Kolmogorov spectrum of phase fluctuations, $\phi(x,y)$. An incident plane wave, $\psi(x,y,h+\Delta h)$, is perturbed by the phase screen after passing through the screen, resulting in a phase shift

$$\psi'(x,y,h) = \psi(x,y,h+\Delta h)\exp\left[i\phi_i(x,y,h)\right], \tag{6.7}$$

where $\phi(x,y,h)$ is the perturbing phase, and Δh is the thickness of the layer.

A simplification of Figure 6.2 is given in a single layer model shown in Figure 6.3. Geometric ray tracing is used in this model to represent the projection of the pupil on a turbulence layer, L, towards a single, NGS. A planar wavefront, $\psi(\cdot)$, is perturbed by L at height $h+\Delta h$. The perturbed wavefront emerges from layer L at height h, and is shown as $\psi'(\cdot)$.

The perturbed wavefront, $\psi'(x,y,h)$, propagates through non-turbulent air to an observational platform, shown in Figure 6.3 as height, h_0. The complex field distribution, $U(\cdot)$, is

related to wavefront phase, $\phi(\cdot)$, by

$$U_h(x,y) = \exp\{j\phi(x,y)\} \tag{6.8}$$

where $\phi(\cdot)$ is the piston removed phase component of $\psi(\cdot)$ as defined by Equation 6.4, and where the relationship between the wavefront and phase is given by

$$\phi(x,y) = \frac{2\pi}{\lambda} W(x,y). \tag{6.9}$$

A geometric model is considered for wavefront propagation through weak turbulence, or for wavefront propagation through moderate to strong turbulence over short paths, where primary or secondary scattering of geometric rays occurring between layers is ignored [49]. Given the absence of diffraction in this simplified geometrical turbulence model, the optical field produced by propagation through layer L_i, where $i \in \{1,2,\cdots,N\}$, can be generalised as $U_{h_i}(x,y)$. The relationship between the complex field at the layer, and the complex field at ground level, $U_{h_0}(x,y)$, is

$$U_{h_0}(x,y) = U_{h_i}(x,y). \tag{6.10}$$

However, for moderate to strong turbulence, the effects of diffraction cannot be ignored. In these cases, Equation 6.10 requires modification, where diffraction between layers is considered, i.e., the complex field at ground level requires convolution with the Fresnel kernel; this is represented as

$$U_0(x,y) = U_{h_i}(x,y) \odot h(x,y), \tag{6.11}$$

where $h(x,y)$ is the Fresnel kernel given as [86]

$$h(x,y) = \frac{\exp(jkh_i)}{j\lambda h_i} \exp\left[\frac{jk}{2h_i}\left(x^2+y^2\right)\right]. \tag{6.12}$$

A practical method to simulate wavefront propagation is given by Johnston *et al.* [64], where additional convolution with a Gaussian smoothing term is performed.

Multiple layer propagation

The geometric model shown in Figure 6.2 depicts a lower, boundary layer, L_1, and an upper layer, L_2; each layer is defined using a different refractive index structure constant, i.e., $C_n^2(h_1)$ and $C_n^2(h_2)$. However, as previously discussed, individual phase screens can be generated for simulation that support a thin-layer model, using separate parameters, r_0 and h.

Given the simplified geometrical model discussed in the previous subsection, the phase shift generated by propagation through L_2, $\phi_2(x',y')$, and subsequent phase shift generated by layer L_1, $\phi_1(x',y')$, can be combined in terms of pupil phase function, $\phi(x,y)$, and generalised as [61]

$$\phi(x,y) = \sum_{i=1}^{N} c_i \, \phi_i(x_i, y_i, h_i), \tag{6.13}$$

where $\phi_i(\cdot)$ is the wavefront phase associated with i^{th} layer at a height h_i, N is the total number of atmospheric layers, and c_i are weights, configured such that

$$\sum_{i=1}^{N} c_i^2 = 1, \tag{6.14}$$

and are used to weight the effect of each layer in proportion to the $r_0^{-5/6}$ theory, thus ensuring normalisation of the accumulated phase, $\phi(x,y)$, in the pupil.

The same field-phase relationship used for single-layer propagation is applied to multi-layer propagation, and is defined by combining Equations 6.8 and 6.11, yielding

$$U_0(x,y) = \exp\{j\phi(x,y)\}. \tag{6.15}$$

The optical field given by Equation 6.15 results in an amplitude intensity distribution, $I(\cdot)$, in the focal plane. Such amplitude intensity distributions are used for both wavefront sensing and image recovery, these topics are discussed in following subsection.

6.2.5 Phase estimation and imaging

The process described in the previous subsection for wavefront propagation is identical for both source beacons and target objects. However, the complex field, representing either an aberrated wavefront or distorted image, as propagation extends from the aperture to the exit

pupil and image plane, are processed differently, according to the type of data required. For example, two slightly defocused images of source beacons define regionalised wavefront maps, whereas a focused image is required for restoration of target objects. Simulating the effects of diffraction, from the aperture to the respective image planes, for either phase estimation of source beacons, or image recovery of a target, is discussed in this subsection.

Phase estimation

The central ray from the aperture to each source beacon, as shown in Figure 6.2, to is used to define the central and regionalised projection of the pupil on each layer. Since phase cannot be measured directly, a WFS is used to acquire intensity distributions at the image plane and convert these intensities into phase estimates.

Propagation of the complex field to the exit pupil, and formulation of two defocused images for wavefront acquisition using the Roddier method for phase extraction [114], is simulated using Fresnel diffraction [19]. The geometric WFS also uses amplitude intensity measurements to estimate phase, however diffraction is not employed [19].

The amplitude intensity at any point in the propagation path is given by [19]

$$I(x,y,z+\Delta z) = \frac{I(x,y,z)}{1+H(x,y,z)\Delta z + K(x,y,z)\Delta z^2}, \tag{6.16}$$

where $H(x,y,z) = \nabla^2 W(x,y,z) = W_{xx}(x,y,z) + W_{yy}(x,y,z)$, the Laplacian or mean curvature of the wavefront and $K(x,y,z) = W_{xx}(x,y,z)\,W_{yy}(x,y,z) - W_{xy}(x,y,z)^2$, the Gaussian curvature of the wavefront [19].

Given the intensity distribution in Equation 6.16, wavefront phase is estimated from two intensity distribution planes, as described in Section 4.4. Phase estimates from both the extended curvature and geometric WFSs are obtained by solving Equation 4.24. The phase estimate, $\hat{\phi}(x,y)$, is expressed in terms of a set of Zernike coefficients, $a_2, a_3, \cdots, a_N$, as defined by Equation 3.18; this is restated here for convenience as

$$\hat{\phi}(r\rho,\theta) \approx \sum_{i=2}^{K} a_i Z_i(\rho,\theta), \tag{6.17}$$

where $R\rho$ is the normalised radial coordinate in the aperture, given an aperture of radius, r.

A range of both photon and Gaussian noise levels are used to distort intensity distributions

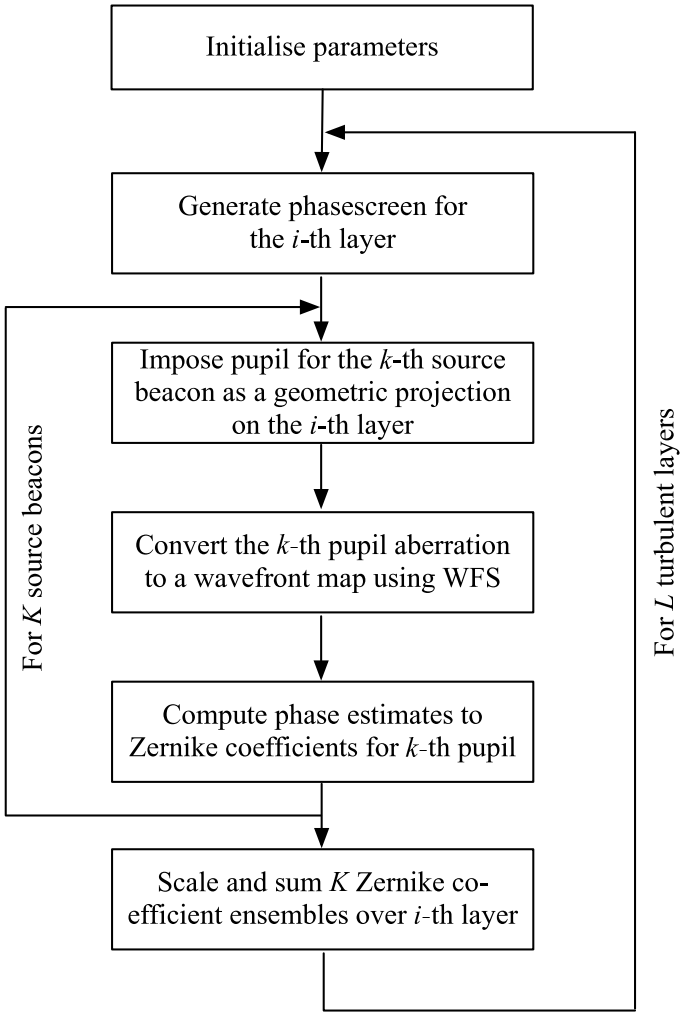

Figure 6.4 Phase screen generation and wavefront phase estimation.

in the pupil, prior to wavefront extraction. As a result, pupil phase functions form a set of random processes, where the noise parameters are altered to assess individual performance characteristics in predicting the anisoplanatic PSF.

Each Zernike set forms a phase estimate, resulting from the wavefront phase perturbation of a specific turbulent layer. The accumulation of each Zernike set representing one layer, is used to define the resulting phase estimate for each source beacon. Simulating these processes, i.e., wavefront propagation, acquisition of corresponding wavefront phase, representation in terms of Zernike coefficients, and lastly, summation to form the total phase estimate of each source beacon in the pupil, is summarised in Figure 6.4.

Given the importance of both spatial and temporal sampling in simulations, consideration

is given to these topics in the next subsection.

Imaging

The effects of wavefront propagation from a target object is considered. The passing of a complex field through a thin lens, such as the objective of a telescope, results in a focused image. Propagation from the entrance pupil to the focal plane is simulated using the Fraunhofer approximation, given by Equation 3.3, where the amplitude transmittance function, $t_A(x,y)$, is defined by Equation 6.3. The resulting intensity distribution is given by [44]

$$I_f(u,v) = \frac{A^2}{\lambda^2 f^2}\left\| \int_{\infty}^{\infty}\int_{\infty}^{\infty} U_{\text{trans}}(x,y)\, \exp[-j\frac{2\pi}{\lambda f}(xu+yv)]dxdy \right\|^2, \tag{6.18}$$

where f is the distance from the exit pupil to the image plane, A is the area of the pupil, and λ is the wavelength. Equation 6.18 is the squared modulus of the PSF given by Equation 3.9, as defined by Equation 3.41 in Section 3.7 for incoherent light.

The resulting target image, shown in Figure 6.1 as $\mathbf{g}_{\text{T}}$, is used by a deconvolution algorithm for restoration. In the absence of turbulence, the resolution is diffraction limited by the optical instrument, i.e., limited to $1.22\lambda/D$, and is represented by the PSF given in Equation 3.41.

According to Goodman [44], the phase distribution is of no consequence to the intensity in the focal plane given by Equation 6.18. However, in terms of astronomical imaging, significant aberrations ($\approx 86\%$) are contributed by tilt, where the x- and y-component of tilt are Zernike polynomials, Z_2 and Z_3. Such phase aberrations displace, rather than distort the PSF. Long exposure images of NGSs highlight this effect as a Gaussian intensity distribution averaged over many iterations, as described by Equation 3.34.

With the application of a suitable deconvolution algorithm, the process used to estimate the original target image, $\hat{\mathbf{f}}_T$, is outlined in Appendix A, using the distorted image, $\mathbf{g}_T$, and the spatially variant PSF. The latter is derived using phase estimates from multiple source beacons, as described in the aforementioned subsection.

6.2.6 Sampling requirements

Simulating the propagation of light through turbulence required the establishment of minimum sampling rates. Since a spatotemporal model was required, both spatial and temporal

sampling rates were defined. These sampling considerations are discussed in the following subsections.

Spatial

The spatial requirements are considered in this subsection. A commonly used criterion to determine angular resolution for optical systems is

$$\Delta\theta = \frac{\lambda}{D}, \tag{6.19}$$

where λ is the optical wavelength and D is the diameter of the aperture.

To determine the minimum spatial resolution, the maximum distance of propagation, z, is also required. The resulting resolution is

$$\Delta x = \frac{\lambda z}{D}. \tag{6.20}$$

Given the simulation parameters from Table 6.1, where the maximum propagation distance is defined as 13000 m, D = 1 m, and the peak wavelength, $\lambda = 0.5 \times 10^{-6}$ m, the spatial sampling in the pupil plane is given as

$$\Delta x \leq 0.0065 \text{ m}, \tag{6.21}$$

and in terms of spatial separation

$$\Delta\theta = 1 \ \mu\text{rad / pixel}. \tag{6.22}$$

Using Equation 6.22, a maximum FOV was calculated to be 470 μradians. In addition, given the range of Fried coherence lengths used in simulations, $0.05 \leq r_0 \leq 0.25$ m, the minimum spatial sampling rate was considered adequate. Lastly, by employing a centrally located science object, angular separations could range from 0 to 235 μradians.

Temporal

The Greenwood frequency, f_G, defined by Equation 4.11, was used to establish the temporal rate. Parameters are required for the propagation simulation environment to establish the temporal sampling frequency, analogous to frames-per-second (FPS) used in CCD camera

software[2] to capture discrete images representing video sequences; related dependencies, in terms of the Greenwood frequency, f_G, are discussed in Section 4.1.3.

Wind is effectively responsible for displacing turbulence over a fixed aperture [49]. According to Taylor's hypothesis, turbulence does not change, i.e, it remains essentially 'frozen' as it moves over the aperture. Thus, phase variations measured from source beacons should remain constant as their effects on phase perturbation pass over the aperture. Given this desirable condition, phase predictions over anisoplanatic regions are validated in this research. This hypothesis is applied to the propagation model used in this study, where phase screens are displaced over the meta-pupil at a velocity, $v(h)$.

The displacement of each phasescreen is directly proportional to the temporal sampling frequency, f_T. Consider a single turbulence layer, such as in the propagation model shown in Figure 6.2 where L_2 has been removed. The observation platform, G, and each of the four projected pupils (disks), remain fixed. However, at each sample period, the model will update L_1 by displacing the contents of the meta-pupil, shown as a large outer disk, to the right for example, and replace the displaced contents with new data from the left. The rate of this displacement is controlled by a wind speed parameter, τ_1, and is proportional to the sampling frequency. Now consider a second layer, L_2, where its direction is orthogonal to L_1, i.e., into the page, and where the speed is controlled by second parameter, τ_2. Thus, at each sample period, phasescreens L_1 and L_2 are displaced using parameters τ_1 and τ_2 in proportion with the sampling period, τ_T. Wind velocity can be expressed as a function of altitude, $v(h)$, and simulated as described.

For simulation of turbulent layers less than 20 km, the Bufton model was considered [103]

$$v(h) = 5 + 30\exp\left\{ - \left[\frac{h - 9400}{4800} \right]^2 \right\}. \tag{6.23}$$

Greenwood [45], proposed a Gaussian model for the average wind profile

$$v(h) = v_G + v_T \exp\left[- \left(\frac{z\cos(\zeta) - H_T}{L_T} \right)^2 \right] \times \left[\sin^2\phi + \cos^2\phi\cos^2\zeta \right]^{1/2}, \tag{6.24}$$

where v_G is the wind velocity at low altitude, v_T is the wind velocity at the tropopause, ζ is the zenith angle of observation, H_T is the height of the tropopause, L_T is the thickness of the tropopause layer, and ϕ is the wind direction relative to the telescope azimuth.

[2]Flycapture CCD application software, Point Grey Research Inc.

However, in applying Equation 6.24, wind velocities at 10 km can exceed 30 ms^{-1}. Parameters based on Equations 6.23 and 6.24 were incorporated in the simulation model. However, given the finite size of the phasescreen when simulating high wind velocities, in addition to the limited propagation distance listed in Table 6.1, the number of samples is limited to 300.

As outlined earlier in this subsection, the Taylor hypothesis is an important prior for predicting new perturbations. In addition, and of equal importance, is the requirement to sample wavefront aberrations from point-source beacons at a sufficiently high rate, in order to maintain a consistent wavefront pattern over the entire field of view. If velocity $V(h)$ is too high, and the sample period τ_T is too low, the resulting time series data used for training and verifying the prediction of new perturbations over the field-of-view would place excessive demands on an ESN predictor, resulting in poor overall performance.

Considering local observational conditions at MJUO, a more appropriate velocity wind model is given by Mohr [91]. In a later analysis using empirical data, this wind profile was employed. A summary of these models is listed in Appendix Table E.1.

6.2.7 Operational summary

Assuming the Taylor hypothesis, and by generating one or more phase screens for multiple wavefront propagation, phase perturbations were extracted from regions on intra- and extra-focal planes, represented by intensity distributions of one or more source beacons. Both curvature and geometric wavefront sensors were used to extract phase perturbations, in terms of Zernike coefficients. The method to simulate Zernike coefficient extraction is shown in Figure 6.5. Ragazzoni *et al.* [110] proposed a similar method to estimate modal aberrations from several source beacons as a practical demonstration of atmospheric tomography. Ragazzoni's method differs from the method described here, where a curvature WFS employing two focal planes was used to reduce scintillation.

Regions representing individual, isoplanatic patches from Figure 6.5 (a) were selected over a single phase screen, and the propagation of broadband light from several source beacons was used to create independent wavefront maps in the pupil plane. The extraction of two and seven isoplanatic patches is shown in Figure 6.5 (b) and(c), respectively. The curvature methods by Roddier [115] and van Dan *et al.* [136] were used to recover wavefront aberrations in the image plane and Zernike coefficients, $a_2, \cdots a_{20}$, from intensity distributions of source beacons S_1, S_2 and $S_1, \cdots, S_7$; this is shown in Figure 6.5 (b) and (c), respectively,

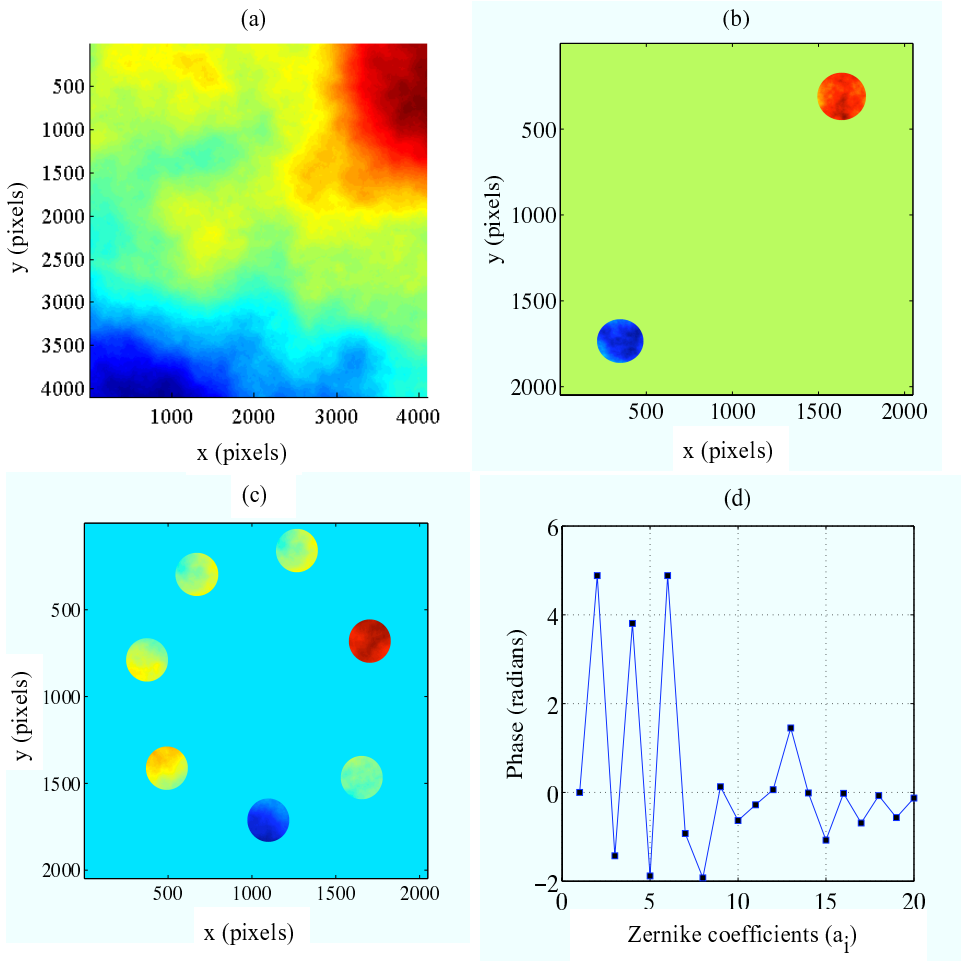

Figure 6.5 (a) Phase screen, (b) multi-aperture mask: 2 reference beacons, (c) multi-aperture mask: 7 reference beacons, (d) Zernike coefficients $\{a_2, a_3, \cdots, a_{20}\}$ from one reference beacon.

for estimation of one Zernike set shown in (d).

Wavefront propagation and corresponding phase estimation was conducted over the visual spectrum to simulate CCD camera[3] performance, as was employed for both the geometric and curvature WFSs. The spectrum used to model each camera was based on the wavelength of light, λ, where $400 \leq \lambda \leq 600$ nm.

6.2.8 Data structures

The propagation model required a data structure that could be employed by the prediction model for estimation and verification. This subsection outlines the data structures used to

[3] model number DX-BW-CSBX

acquire and store data resulting from propagation simulations, before processing by the second simulation phase using the predictive model.

The predictive model requires specific data structures for input to the recurrent neural network. An identical data format described here is used for field data acquisition in the laboratory and for observation runs. However, in terms of data processing, significant differences between field and simulated modes of operation are:

1. Image data from two CCD cameras replaced the propagation model described in Subsection 6.2.4.
2. Corrective rotations due to CCD orientations were required for extraction of intra- and extra-focal image planes.
3. Post processing was required to extract each pair of source and target objects, after defining respective regions of interest (ROI).

The data structure used to store training samples is shown in Figure 6.6. Data ensembles used for training the ESN were generated as a result of performing simulations on the propagation model. The structure comprises random data sequences of Zernike coefficients, $\mathbf{z}$; these form time-series training sequences, $\mathbf{Z}_n$, where n is the n^{th} training sample, $n \in \{1,2,\cdots,N\}$, from N time steps. Each training ensemble, r, was randomly selected from locations in the metapupil.

The training data structure shown in Figure 6.6 can be represented as a 2D data structure. This is shown as two, horizontal data blocks in the lower portion of Figure 6.7, labelled *Output data*. Each series of blocks represents an ESN output training ensemble for one source beacon; additional training sample blocks for K sources are labelled, *Sources*, $S_1,\cdots,S_K$.

The second set of data structures, shown in the upper portion of Figure 6.7, is labelled *Input data*. This data structure is composed of two sets, a Zernike coefficient matrix, $\mathbf{Z}$, as defined by Equation 5.6, and the separation matrix, $\boldsymbol{\theta}_{\text{sep}}$ as defined by Equation 5.7.

To demonstrate how ESN training data are stored, consider two training samples, shown as τ_1 and τ_2 in Figure 6.6. Both datum represent a Zernike coefficient, a_m, assigned to two training ensembles, P and Q, for source beacon, S_1, and at time-sample, n. Each datum is shown in corresponding 2D matrices as τ_1 and τ_2 in Figure 6.7. Collectively, data ensembles *Input data* and *Output data* were generated from wavefront propagation simulations over N

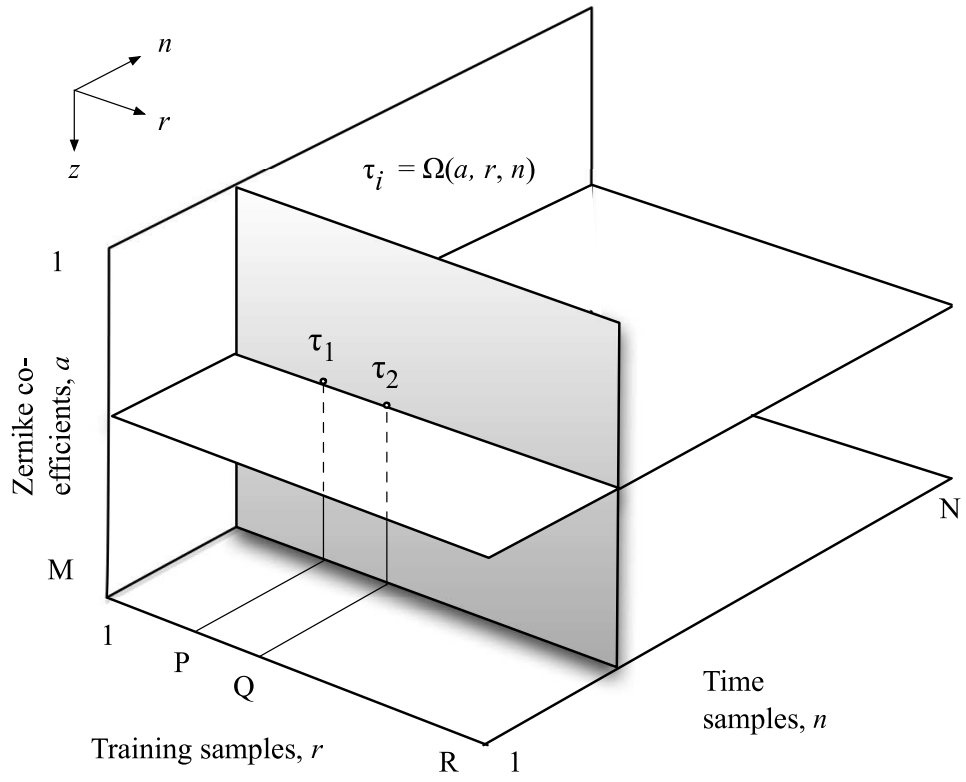

Figure 6.6 Training samples data structure.

time samples. Each ensemble was used correspondingly to train an ESN for prediction of Zernike coefficients of a target object.

Given a series of time sampled data, $\tau = \{1 \cdots M\}$, comprising a set of Zernike coefficients, $a_1, \cdots, a_M$, and source to target separation data, θ, where $\theta_{max} \geq \theta \geq 0$, for each source object, $S_1, \cdots, S_K$, the total number of ESN inputs, U, for prediction of a target T over region $\mathbf{R}$ is

$$U = N(K+1). \tag{6.25}$$

For example, if three source objects are used to predict the first 20 Zernike coefficients perturbing the image of a target object within region $\mathbf{R}$, the size, U, of ESN input vector, $\mathbf{u}$, is $3 \times (20+1) = 63$ input data signals.

Based on the large number of inputs required, and in order to reduce the complexity of the ESN reservoir (and address possible performance issues), the number of coefficients was reduced to two, normalised signals. Thus, a modular ESN structure was used for prediction,

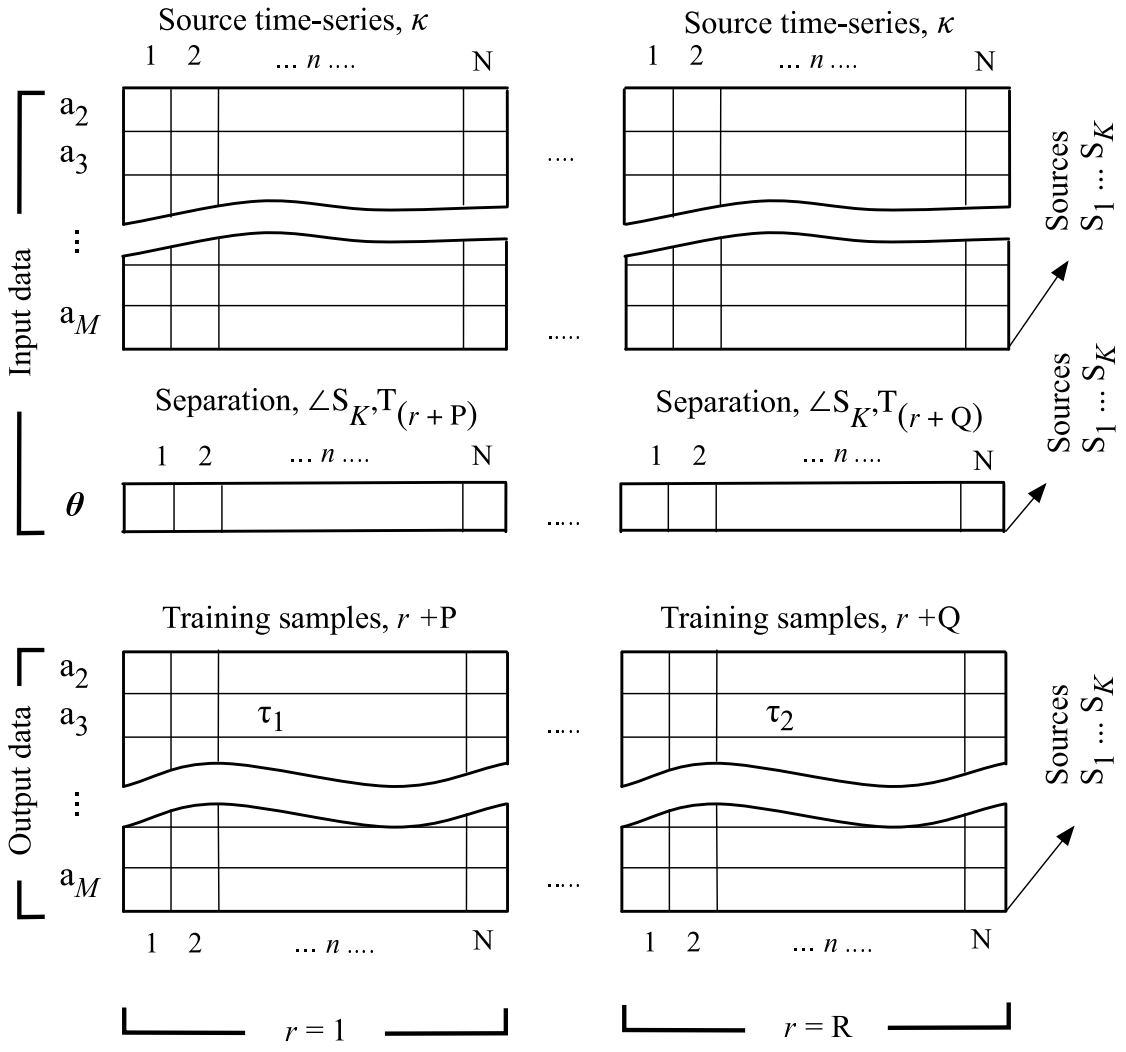

Figure 6.7 Propagation data structures.

where 10 pairs of Zernike modes were combined to estimate a residual wavefront error, ε^2_{20}, of a target object, T.

6.3 The prediction model

Various recurrent neural network architectures have been used in the field of adaptive optics for prediction [90]. However, such architectures have primarily been used for temporal compensation of servo control systems [94] due to decorrelation of the atmosphere [76]. In this section, a predictive architecture, such as detailed in Section 5.3, is extended for the

Table 6.2 Prediction Model Parameters.

Prediction Parameter	Values		
	Min.	**Opt.**	**Max.**
Number of inputs, U	6	18	96
Number of outputs, Y	1	2	11
Spectral radius, $\rho(\mathbf{W}_{\mathrm{DR}})$	0.1	0.95	0.99
Network sparseness, $0.5 = 50\%$ sparsity	0.1	0.1	0.99
Number of hidden nodes, W	100	600	1200

estimation of Zernike coefficients over anisoplanatic regions. To achieve this, a training algorithm was developed to provide spatiotemporal capabilities, and a modular framework was implemented.

The framework facilitated the spatiotemporal training of the ESN using system parameters and data ensembled from the propagation model, and establishes a systematic process for testing and verifying network performance. Lastly, a verification model of the ESN predictor is presented.

6.3.1 Prediction model parameters

Justification for a set of system parameters is twofold. Firstly, the ESN architecture uses a set of parameters for optimisation. A process of *tuning* these parameters to establish optimal computability in terms of network dynamics and data-sets presented to the network, is required. Secondly, a set of system parameters defined a range, within which each parameter could be assigned, and optimal values stored for repeated application.

A summary of system parameters used in the prediction model is listed in Table 6.2. A detailed analysis of network optimisation is presented in Section 5.4.

6.3.2 Spatiotemporal training

Training the RNN was based on spatiotemporal data from several source beacons and a single, 'floating' target. Data ensembles representing M Zernike coefficients from K

source beacons were acquired over N time-steps. These data ensembles represented a set of Zernike coefficient vectors (ZCVs). Additionally, separation data, represented by a set of spatial training matrices (STMs) were generated; both data sets were used to train an ESN to predict a target perturbation. Methods developed to train ESNs over both single, and multiple layers, are described in this subsection.

Single layer

Phase fluctuations in the pupil, caused by the effects of atmospheric turbulence over the optical paths of three, widely separated source objects, S_1, S_2, and S_3, can be converted into wavefront maps, each representing individual phase distortions. The centre of each wavefront map is located at a specific coordinate within the metapupil. During simulations, these coordinates remained fixed, as would be the case for NGSs or other source beacons. Figure 6.8 shows each wavefront map as the projection of the pupil, separated from the origin, O, by vectors, $\mathbf{s}_1$, $\mathbf{s}_2$, and $\mathbf{s}_3$.

The ESN was trained with R wavefront maps over the metapupil; this region is marked A_{T} in Figure 6.8. Given a trained network, the prediction of wavefront maps corresponding to a target, using input perturbation sequences from several source beacons, was proposed. The training of the network using angular separations was a crucial requirement to allow selection of any arbitrary target location over the metapupil.

As required for ESN training, a set of angular separations between each source and a possible target was generated using a normally distributed, pseudo-random sequence. Each set of separations required a Cartesian coordinate to locate the centre of the target wavefront map to a corresponding position on the phase screen. For example, the centre of a possible target used in training is shown in Figure 6.8 as $T(x,y)$.

The procedure used to find the centre of each training wavefront map is summarised as follows. A minimum of three vectors is required, where each vector is associated with a source beacon. These are shown in Figure 6.8 as $\mathbf{q}_1$, $\mathbf{q}_2$, and $\mathbf{q}_3$. The length of each vector was found by

$$q_k = h\,\xi\,\tan(\theta_k), \tag{6.26}$$

where h is the height of the turbulent layer, θ_k is the angular separation between source, S_k, and a target, $T(x,y)$, for $k \in \{1,\cdots,K\}$, and ξ is a scaling factor used to meet the spatial sampling requirements defined in Subsection 6.2.6.

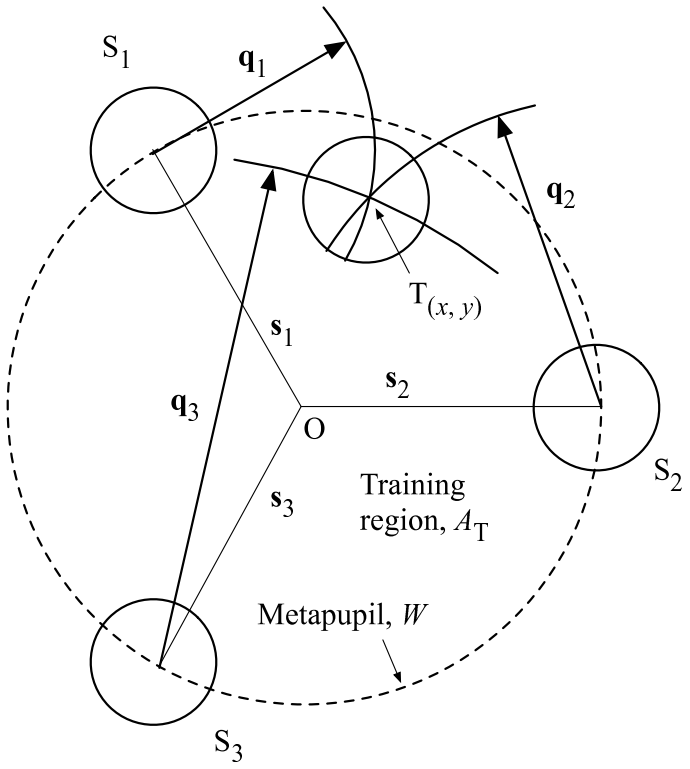

Figure 6.8 Angular separations used for referencing training locations.

With reference to Figure 6.8, the projection of the pupil on to a turbulent layer is similar to "stretching" each source and target vector; i.e., as the height of the turbulent layer increases, each wavefront map and corresponding vector increases in length. If the turbulence were directly over the pupil e.g., $h = 0$, the length of each vector would be zero.

It was possible to find a target location by solving the intersection of K arcs, generated by the rotation of vectors of known length, and where the end point of each vector is fixed at the centre of a corresponding wavefront map. Mathematically, this can be represented as

$$\phi(x,y) = \cap_{k \in K} P_k, \tag{6.27}$$

where $\phi_k(x,y)$ is phase aberration at the centre of the training wavefront map $T_k(x,y)$ from phase screen $\boldsymbol{\phi}$, and P_k is the (x,y) coordinate for the k^{th} source, given as,

$$P_k = \begin{cases} x_k = a + q_k \cos(\theta_k) \\ y_k = b + q_k \sin(\theta_k), \end{cases} \tag{6.28}$$

and where a and b are the rectangular coordinates of the centre of the pupil of source beacon S_k and projected on turbulence layer L, q_k is the length component of displacement vector, $\mathbf{q}_k$, given by Equation 6.26.

Subsequently, Equation 6.28 was used to generate the separation sequence data used to train the ESN.

Multiple layers

The multiple layer model shown in Figure 6.2, can be described as follows. Training vectors were calculated for each layer, where the position of each training vector was corrected for the height of each turbulent layer, h_i, where $i = 1, 2, \cdots L$ layers of turbulence. This is shown in Figure 6.9, where offset positions over the metapupil, and in terms of angular separations, are shown. For example, the new source vector for source S_1 at layer L', is shown as S_1'. In simulations propagating wavefronts from sources S_1 to S_3, phase perturbations were offset for each layer, as a function of the height of the turbulence, Z_n. As the perturbations from each layer were summed for each time-step, as given by Equation 6.13, the location from each layer formed an offset from the centre of the pupil. In addition, the training region was expanded to allow for corresponding positions of training samples from higher layers. However, since the wavefront is considered effectively planar, the size of each subpupil projected onto each layer remains the same. This method is consistent to that shown by Ragazzoni [108] for multiple layers, and is distinct from the method used for spherical wavefronts as generated from laser guide stars.

An example of a 2-layer model is shown in Figure 6.9. In this example, training samples, in terms of phase perturbations from $T(x,y)$ at layer L_1, were added to phase perturbations at $T'(x,y)$ and taken from layer L'. In this regard, an accurate formulation of phase distribution between layers was achieved.

Spatial coordinates for each training vector were selected over a phase screen using Gaussian statistics over a circular region of radius 996 pixels. In terms of angular separation of a target at zenith to a source beacon, i.e., $\angle T, S_n$, the range used was between 0 to 242 μradians; in terms of angular orientation, 0 to 360 Degrees was measured from East. A histogram, based on the normal distribution used to generate the training vectors through angular separation, $0 \leq \theta_T \leq 240\,\mu$rad, and angular orientation, $0 \leq \phi \leq 360$ Degrees, is shown in Figure 6.10.

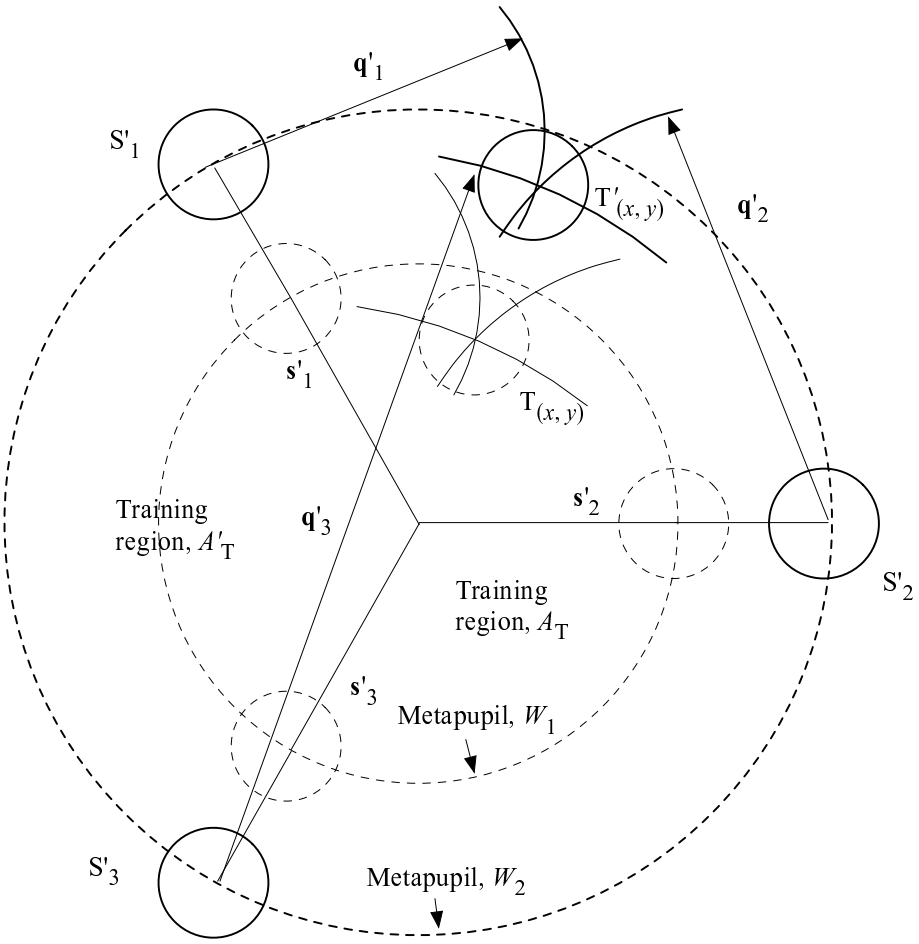

Figure 6.9 Formulation used for calculation of angular separations for two turbulent layers.

6.3.3 Data generation algorithms

The simulation framework supporting the propagation of multiple point-sources through turbulence, and representation of each source in terms of spatial and temporal domains was devised to generate random data ensembles. Each set of propagation simulations generated two data ensembles, and each can be described in terms of an algorithm. Both algorithms are similar in content, however the sequencing of key operations within each process for generating either training or test data are significantly different.

The algorithms listed in this section are the result of simulating the effects of propagating K bright, point-source beacons and multiple targets of variable magnitude and position, using an imaging system and a simple prime focus telescope, supporting a 1-m aperture and focal ratio of f 13.5. A variety of configuration, environmental, and physical parameters were provided in each simulation, and these are listed in Table 6.1.

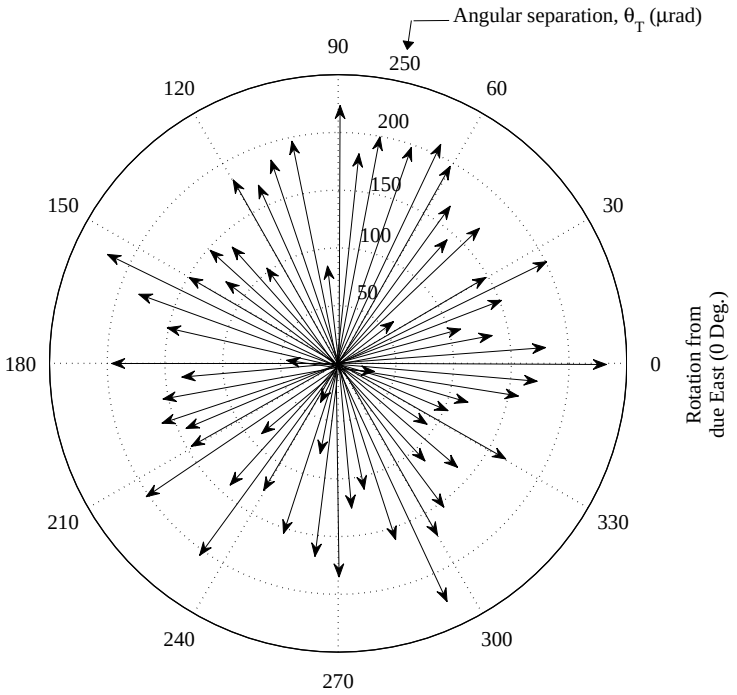

Figure 6.10 Set of 60 angular separation training vectors where the length of each vector is generated using the normal distribution.

The collection of time-series training data from multiple source beacons was achieved using the algorithm shown in Algorithm 1. One or more phase screens were instantiated and moved over a fixed aperture by winds of velocity, v_L, each with a unique bearing of either, North, South, West, or East. Fresnel diffraction was used to image two slightly defocused, intra-focal and extra-focal images, comprising multiple, point-source objects, aberrated by propagation through turbulence. This is performed in the *getFOV* routine. The region of interest (ROI) surrounding each slightly defocused pair of source images is extracted in routine *getSubRegions(.)*, and aberrations, in terms of Zernike coefficients, are converted using both curvature and geometric sensors in *getZTerms(.)*.

Time-series target data ensembles are generated from Algorithm 2. These ensembles comprise actual target data, in terms of Zernike coefficients, from anisoplanatic regions, and are used to verify and assess network performance.

The resulting data ensembles comprised modal aberrations in terms of Zernike coefficients and angular separation data. These ensembles were saved for further processing by the ESN

```
initialisation;
layer ⇐ 1;
realign (phasescreen);
for sample ⇐ 1 to MaxSamples do
    for timestep ⇐ 1 to MaxTimesteps do
        tLoc ⇐ pRandomLoc (srcSep, apSize);
        ESeps ⇐ getSeparations (tLoc);
        for layer ⇐ 1 to MaxLayers do
            iMask ⇐ getFOV (tLoc, phasescreen);
            subRegions ⇐ getSubRegions (iMask);
            [Zgs] ⇐ getZTerms (subRegions);
            updateLayerLoc (tLoc, z (layer));
            if layers < 1 then
                [Zcs, Zgs] ⇐ getZTerms (subRegions, Zcs, Zgs);
            end
        end
    end
end
```

Algorithm 1: Training data generation algorithm.

```
for target ⇐ 1 to MaxTargets do
    tLoc ⇐ pRandomLoc (srcSep, apSize);
    ESeps ⇐ getSeparations (tLoc);
    PSep2org ⇐ updateLayerOffsets (layer);
    for layer ⇐ 1 to MaxLayers do
        rollbackPhasescreens ( phasescreens(1 ··· N));
        dis ⇐ getRelativeLoc (tLoc);
        for timestep ⇐ 1 to MaxTimesteps do
            tSubRegion ⇐ getTargetRegion(dis);
            [Zcs, Zgs] ⇐ getZTerms (tsubRegion);
            storeNTargetZerns (Zcs, Zgs);
            tLoc ⇐ movePhasescreen (phasescreen, veloc, offset);
        end
        offset ⇐ updateLayerOffsets (PSep2org);
    end
    [Zcs, Zgs] ⇐ sumZerns (Zcs, Zgs);
    storeTarget (Zcs, Zgs, ESeps);
end
```

Algorithm 2: Target data generation algorithm.

for training, prediction, and verification.

6.3.4 Spatiotemporal prediction algorithm

Data acquired from propagation simulations were used to predict Zernike coefficients, representing the SVPSF, over N timesteps. This section outlines a new spatiotemporal prediction algorithm that was developed to firstly predict, and secondly to reconstruct, the SVPSF. The essential stages include:

1. Data preprocessing: data required serialising and normalisation for presentation to the network.
2. Network instantiation: formulation of the network using fundamental parameters, such as defining the number of network nodes.
3. Training: training data collected during the propagation stages were used to train the network.
4. Verification: both target and predicted data were compared using the MSE metric.

Data preprocessing was required to reorder and normalise data for application to the predictive algorithm. The preprocessing operations included:

- Sequencing: training data were serialised for presentation to the network. This was required for the spatotemporal training algorithm to process repeated sequences of training data, acquired during propagation simulations.
- Normalisation: Each Zernike coefficient and spatial data ensemble required normalisation for presentation to the network. This allowed the maximum effectiveness of the network to be applied to each ensemble. In addition, both training data and verification data were normalised. The parameters that allowed renormalisation of data for reconstruction of the SVPSF were preserved, in accord with a layer scaling factor given by Equation 6.13.

Three random ensembles were collected and comprised noise-induced, time-series perturbation data from K point-source beacons, training perturbation data from R training locations, and verification data from J target objects and target locations. Spatial data, in

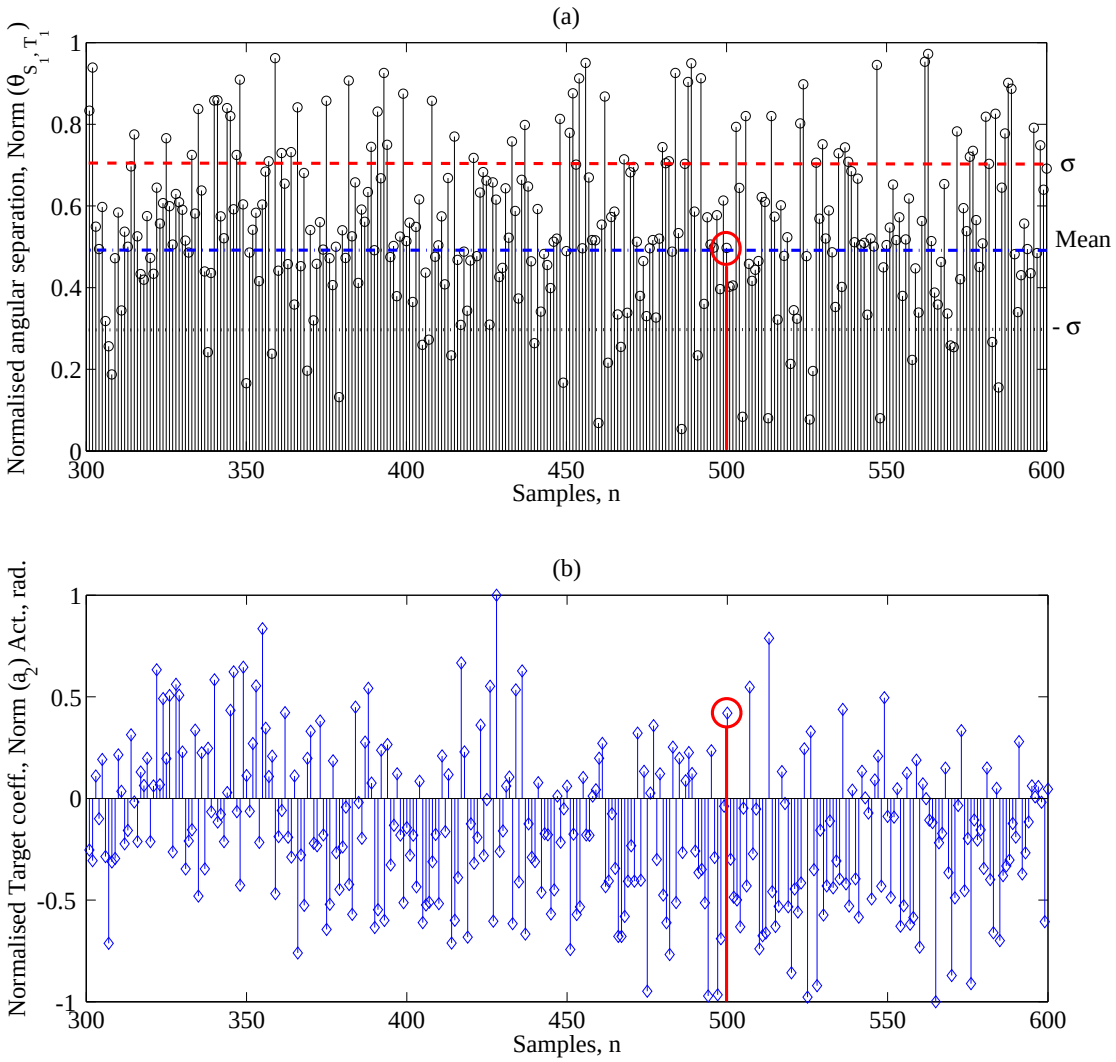

Figure 6.11 Spatiotemporal training data, (a) Normalised angular separations of target objects from reference beacons, θ_{S_1,T_1}, where each data-point represents a training vector; (b) Zernike coefficient data (Actual) from target object T_1, a_2, at angular separations given in (a). An example of a strong correlation between a training vector and actual coefficient value, a_2, is shown at sample point 500.

terms of angular separations from each reference beacon to a single target were collected using the technique outlined in Subsection 6.3.3. Figure 6.11 shows an extract of 3000 training ensembles. The normalised angular separation data in Subfigure (a) was generated using a pseudo-random sequence for a target training sequence; a normal distribution was used. Subfigure (b) shows a series of samples for a target, where a correlated sample is highlighted for Zernike coefficient, a_2.

As previously described, K source beacons were evenly spaced over a wide FOV and remained fixed near a boundary within the metapupil, W_j. The boundary condition was dependent on the altitude of the turbulent layer, h, and as stated previously, length vector shown in Figure 6.9 as $\mathbf{q}'_k$, will increase in proportion with h.

An angular separation between source beacons, $S_1, \cdots, S_K$, and a target object, T, defined as, $\theta_{S_1,T} \cdots \theta_{S_K,T}$, was obtained using a normal distributed, pseudo random number sequence. Each angular separation was combined with Zernike coefficient data, formed by the series accumulation of perturbations caused by propagating each source through K layers of turbulence. The volume of this region is formulated using 3D spatial coordinates, where the planar parameters was stored using polar coordinates, $(\theta_{S_n,T}, r_n)$, and the h parameter (altitude at zenith) comprises Zernike data from each layer.

Using the geometrical model outlined in Subsection 6.2.1, and the accumulation of Zernike coefficients as the basis for multi-level wavefront propagation [49], the projection of the pupil towards each source provides a set of K coordinates that correspond to the height of each turbulent layer. Using this height, and the angular separations from the target to each source, the pupil regions over the phase screen that define the projection of the pupil toward each source, are calculated. By employing the thin layer, geometric propagation model outlined in Section 6.2, phase perturbations in the form of Zernike coefficients are calculated for each layer and form a partial sum. Each Zernike coefficient is then scaled in accordance with Equation 6.13, and normalised to $[1, -1]$ for presentation to the ESN. As each phase screen (per layer) is moved with respect to a fixed aperture, the set of inputs described above form time-series data sets. During propagation, two algorithms, photon and CCD read noise, are used as the basis to convert each time-series data set to random ensembles.

The ESN is then trained using time-series random ensembles from each source as input, and training samples taken from training regions, A_T^i, where $i = 1 \cdots L$ layers; this is shown as *IData* and *TData* data ensembles in Algorithm 3. Both data ensembles are used to construct the state vector, $\mathbf{x}(n)$, as shown in Figure 5.1. Thus, data ensembles generated by the algorithm described in Algorithm 1, is employed to complete the training of the network. The trained ESN network is then used to firstly analyse the eigenspace for network optimisation, and secondly to verify the network using verification data. These procedures as shown Algorithm 3.

```
ZernData ⇐ getZerns (ZernTest, ZernTrain, ZernSrc, SepsTest, Seps);
NormZernData ⇐ Normalise (ZernData);
ProcNZernData ⇐ encodeZernData (NormZernData);
IData ⇐ ProcNZernData.SrcData;
TData ⇐ ProcNZernData.TrainData;
for target ⇐ 1 to MaxTargets do
    ActTargetData ⇐ ProcessZernData[target].Actual;
    TestSrcData ⇐ ProcessZernData[target].Source;
    for average ⇐ 1 to MaxAverage do
        net ⇐ newESN(ESNparameters);
        TrainedNet ⇐ trainESN(IData, TData);
        AnalysedNet[average] ⇐ analyseNet(TrainedNet, ActTargetData, TestSrcData);
        Results[average] ⇐ verifyNet(TrainedNet, ActTargetData, TestSrcData);
    end
end
ProcessedResults ⇐ ProcessResults (Results);
```

Algorithm 3: Zernike prediction algorithm.

Separation data from the propagation simulations are normalised prior to training and verification. Each separation represents an angular displacement from each source to a target of 0 to 242 μrad; the method used for this procedure is shown in Figure 6.11. As outlined above, the training algorithm uses a normal distribution however a mean of half the diameter of each circular training region is imposed. The region defined by $\pm\sigma$ and the mean, `Mean`, is shown in Figure 6.11 (a). As discussed, each training region is defined as the circular region that encompasses the projection of the pupil, forming a metapupil, and is projected on to each turbulent layer. For example, the training region for layer L_1 is shown as A_T in Figure 6.8, whereas A'_T represents the training region for a higher altitude layer, shown in Figure 6.9.

Once a training location for layer L_i is determined, the Euclidian distance that separates each source beacon, $S^i_1, \cdots, S^i_K$, to $T^i(x,y)$ is calculated. These are shown as vectors $\mathbf{q}'_1$, $\mathbf{q}'_2$, and $\mathbf{q}'_3$ in Figure 6.9. The final step before normalisation is converting each vector to an angular separation from each source beacon, by the procedure described in Section 6.3. A geometrical view of the configuration for a 2-layer model is shown in Figure 6.2, where $\theta_{\gamma\varepsilon}$, $\theta_{\beta\varepsilon}$, and $\theta_{\alpha\varepsilon}$ are the angular separations from target ε, within a field-of-view shown as θ_{FOV}.

6.3.5 Verification model

The system model shown in Figure 6.1 integrated the ESN architecture shown in Figure 5.1, and the performance of the overall system was verified using a verification ensemble collected during each propagation simulation run. The verification ensemble is shown in Algorithm 3 as *ActTargetData* and comprised Zernike coefficient data taken from J target regions. Each time-series dataset, comprising one target over the FOV, was verified using the procedure, *verifyNET*$(\cdot)$. Input data ensembles used to train the network are shown in Algorithm 3 as *TestSrcData*.

Figure 6.12 represents the verification model used to determine the MSE metric and assess system performance. Shown are input Zernike coefficient matrices, $\mathbf{Z}_1 \cdots \mathbf{Z}_K$, where $K \geq 3$ source beacons, and where $\mathbf{Z}_k$ is defined by Equation 5.6. Also shown is the separation matrix, $\boldsymbol{\theta}_{\text{Sep}}$, as defined by Equation 5.7. Estimated output data from the network, $\hat{\mathbf{Z}}_{\mathbf{T}}$, for targets, $\mathbf{T} \in \{T_1 \cdots T_J\}$, are compared with actual Zernike coefficient data, $\mathbf{Z}_{\mathbf{T}}$, to calculate the MSE.

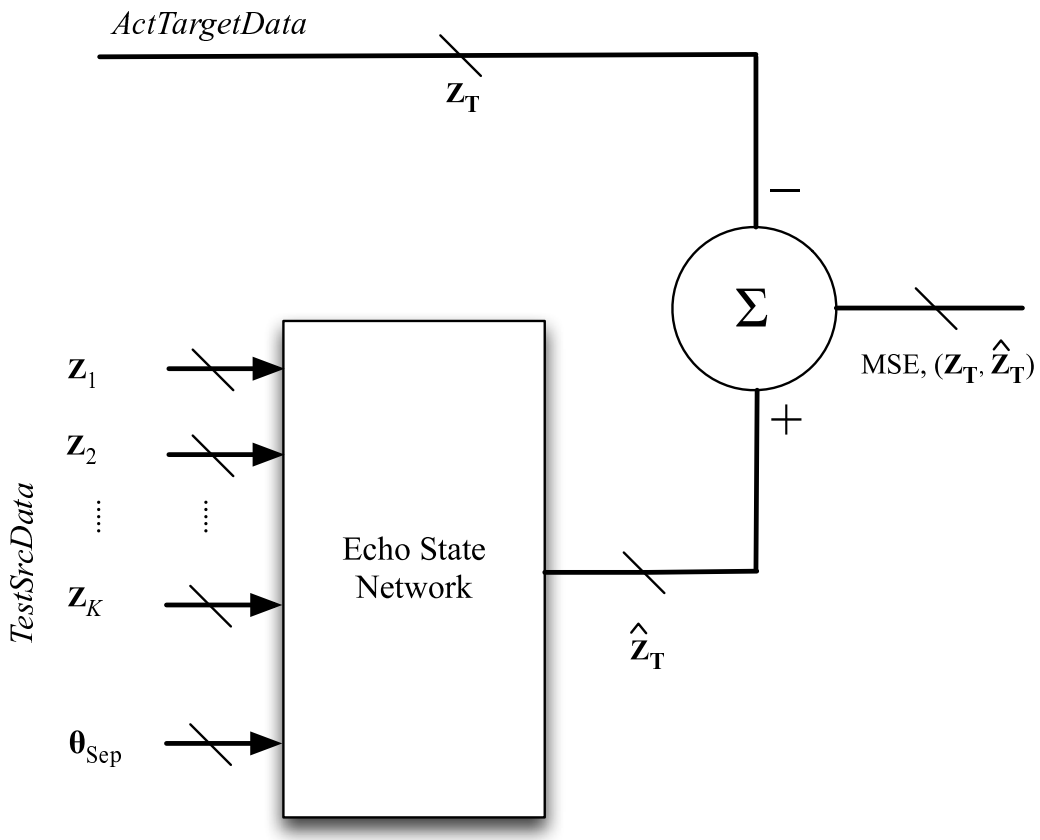

Figure 6.12 Verification model for obtaining the MSE metric.

Chapter 7

The ESN Method

In this chapter a performance analysis is presented, specifically concerning the ESN method outlined in Chapter 1. The simulation framework discussed in Chapter 6 provided data for this analysis. Given the environmental conditions and instrumentation used to acquire astronomical data, random variables and signals were generated, using both Gaussian and Poisson distributions.

Specifically, random variables, such as Zernike coefficients, were used to assess the performance of the ESN method. Low-order Zernike modes were predicted by an ESN, using several reference beacons in the anisoplanatic region. Given accurate predictions of Zernike coefficients, the anisoplanatic PSF can be recovered, thus providing a significant contribution to astronomical image restoration.

The following section describes the general test procedure used in this analysis, and parameter ranges for wavefront propagation simulations. Field angle results are presented in Section 7.2, where angular separations over anisoplanatic regions are varied, and predicted target values are compared with actuals. Section 7.3 evaluates network performance over relatively small field angles, as would apply, for example, in the isoplanatic region. Section 7.4 evaluates ESN performance in terms of Fried's atmospheric coherence length, r_0. This is followed by Sections 7.5 and 7.6 where an analysis on the effects of CCD read and photon noise is presented, respectively. Combinations of these noise conditions are discussed in Section 7.7. The chapter concludes with a performance summary of the ESN method in Section 7.8.

7.1 General test procedure and parameter sets

The data ensembles used in this chapter for analysis were generated by performing two separate simulations. The first used the wavefront propagation model, based on a set of parameters to generate data ensembles described in the previous chapter. The second simulation used an ESN architecture, where the parameters for an optimised ESN remained constant throughout these tests. Optimised ESN values are listed in Table 6.2.

Wavefront propagation parameters were varied according to the required test. For example, several simulations required a fixed angular separation and variable noise parameters, others used the reverse. In all cases, a single target was defined at the centre of the metapupil, shown as O in Figure 6.8, and a single perturbing layer was generated for each simulation. The perturbing layer was moved with a uniform velocity of 5 $\mathrm{m\,s^{-1}}$across a projected pupil plane, at height, h. Simulation parameters used for wavefront propagation are listed in Table 7.1.

Table 7.1 Propagation simulation parameters.

Simulation Parameter	**Values**	
	Min.	**Max.**
Photons, N_p	150	6.23×10^5
SNR, (dB)	80	20
Propagation time-steps, N	300	500
Source beacons, K	3	8
Predicted Zernike modes, M	2	11
Propagation layers	1	2
Turbulence height, h, 10^3 m	5	12
Wind velocity, $V_{\mathrm{Wind}}(h)$, $\mathrm{m\,s^{-1}}$	3	5
Fried Seeing Parameter, r_0, m	0.05	0.5
Training targets per time-step	7	10

Lastly, the range of values in Table 7.1 is based on empirical evidence [91]. For example,

irradiance measurements from focal plane images at MJUO provided insight into the values used in simulations.

7.2 Prediction by field angle and sources

In this section the results from wavefront prediction using variable angular separations are presented. Three sets of results are shown, where each set employs one of three groups of source beacons for prediction.

Firstly, ensembles comprised 1000 sets of Zernike coefficients were formatted and used to train an ESN predictor. Each resulting time-series represented random data, however minimal noise constraints were imposed, simulating ideal irradiance conditions from K source beacons. Thus, each simulation employed 80 dB read noise and 1.6275×10^5 photons.

Figure 7.1 represents the AMSE prediction error of individual Zernike coefficients, $a_2, \cdots, a_{11}$, where the angular separation between a single target and 3, 6, and 8 source beacons was varied between 24 μrad and 242 μrad. A separate phase screen was generated for each simulation, where turbulence strength was defined using the Fried parameter, $D/r_0 = 10$. The first four Zernike coefficients, $a_2, a_3, \cdots, a_5$, corresponding to x-tilt, y-tilt, defocus, x-astigmatism, and y-astigmatism, are marked in Subfigures (a), (c), and (e).

The results shown in Figure 7.1 can be interpreted as follows. Firstly, as the field angle is progressively reduced, a corresponding improvement in predicting target aberrations, in terms of normalised AMSE in (a), (c), and (e), is evident. This improvement applies to the set of Zernike coefficients acquired in training, $\mathbf{z}_\mathrm{T}$, where $\mathbf{z}_\mathrm{T} = \{a_2 \cdots, a_{11}\}$, and can be represented in terms of reducing the area of the metapupil as shown in Figure 6.8. With reference to this figure, as the angular separation between each source and target is reduced, training vectors $\mathbf{q}_k$, the size of the metapupil, W_T, and training area, A_T, are also reduced. Since the number of training samples used for each simulation was constant, the ESN learned spatiotemporal responses over a reduced area, thereby improving the capability of the network in predicting aberrations from a target, $T(x,y)$, within region, A_T.

Secondly, the responses of individual Zernike coefficients, shown in Subfigures (b), (d), and (f) in Figure 7.1, were considered. Considerable variance was noted for each separation over the set of Zernike coefficients, where the range for 8 source beacons used for prediction, can exceed 30%. However, variance is minimised in Subfigure (e), for 24 μrad. Similar examples are shown for smaller field angles discussed in Section 7.3. In addition,

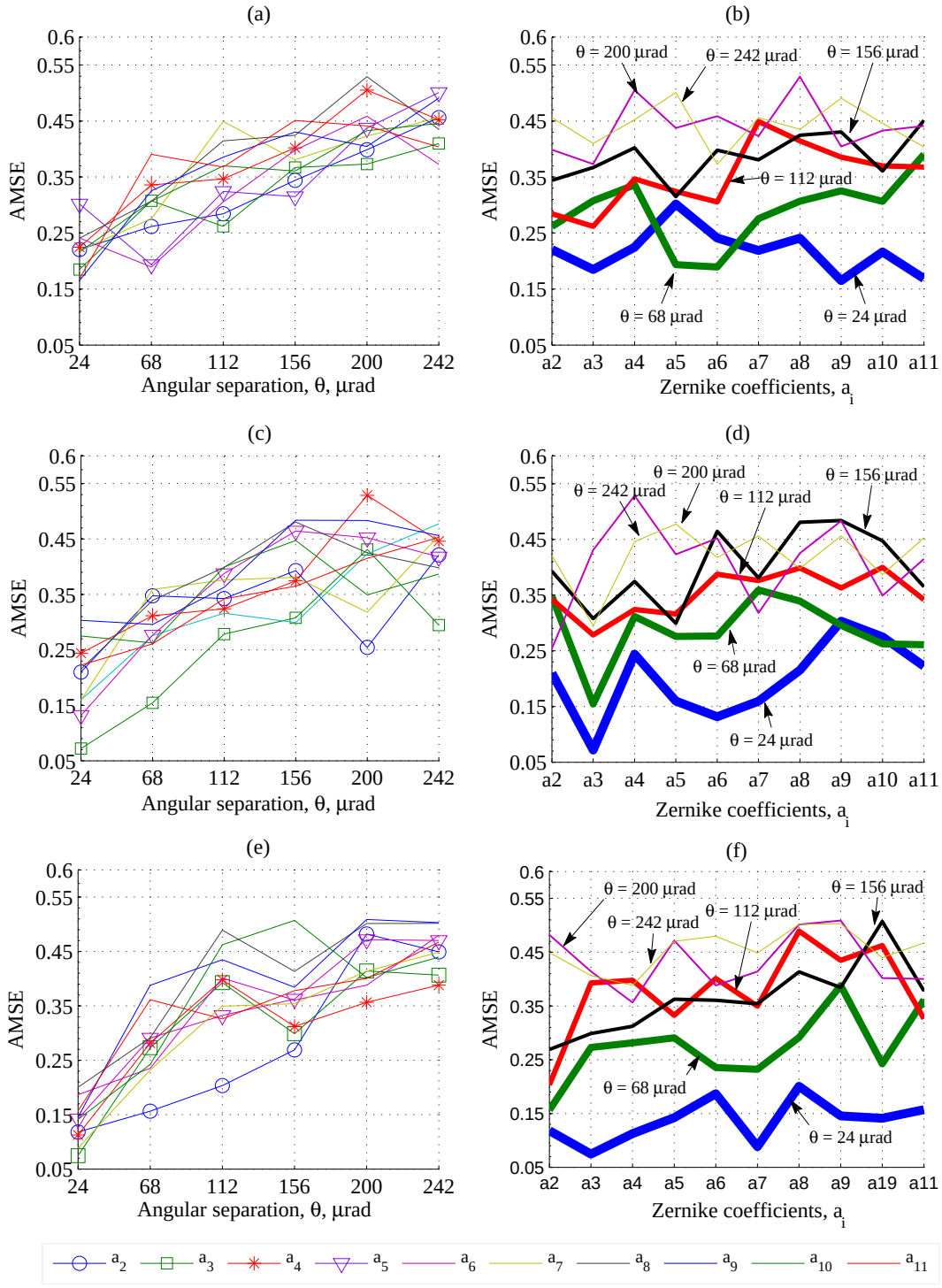

Figure 7.1 First 11 Zernike coefficients average mean squared prediction error by angular separation, θ, and source beacons; (a) & (b) 3 source beacons; (c) & (d) 6 source beacons; (e) & (f) 8 source beacons. All simulation plots were generated using 1000 time samples, 1.6275×10^5 photons, 80dB Gaussian noise, over various source-target separations as shown.

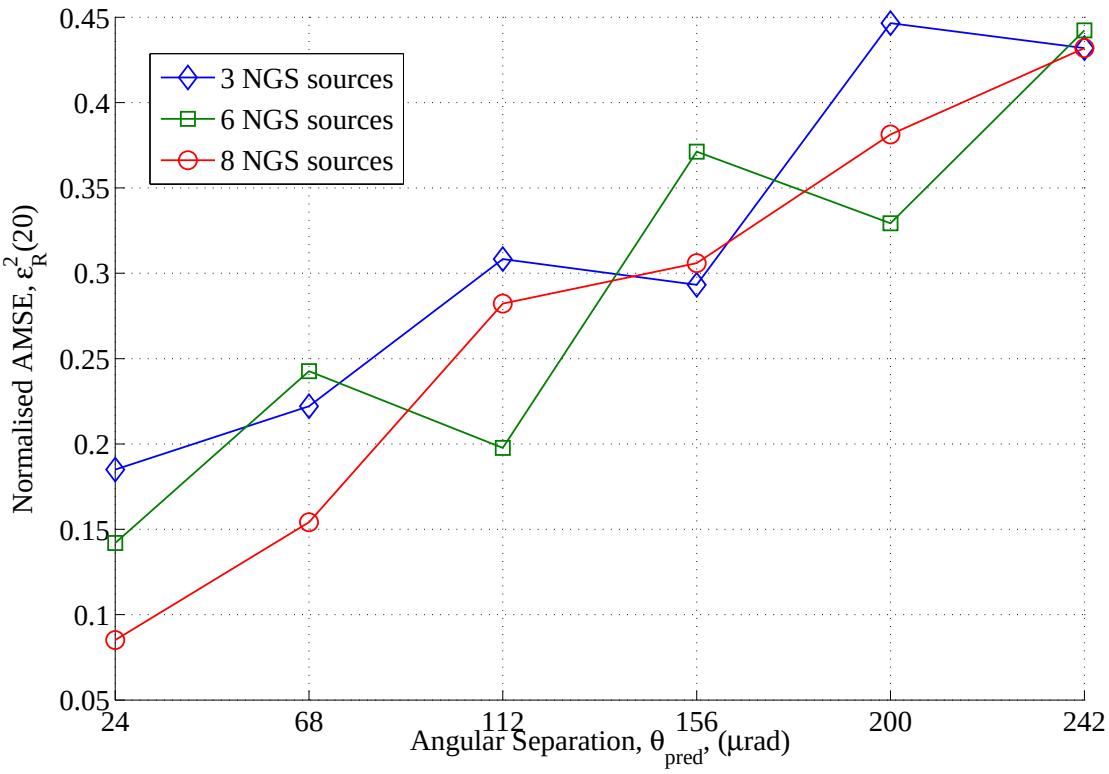

Figure 7.2 Average mean squared error (AMSE) of prediction by angular separation, θ, generated using the geometric wavefront sensor, phase screen with $D/r_0 = 10$, 1.6275×10^5 photons, 80 dB Gaussian noise, and separation, θ, of target and equally spaced sources, ranging from 24 to 242 μrad.

the four lowest modes, highlighted with icons on the legend, tend to dominate the group, in terms of contributing to the lowest AMSE. Subsequent simulations have shown that, given minimal noise constraints, lower-order Zernike modes typically resulted in lower AMSEs, compared to higher-order modes.

Figure 7.2 represents the results shown in Figure 7.1 after renormalisation and accumulation of the set of Zernike terms described earlier in this section; the target wavefront error is shown for 3, 6, and 8 source beacons. Each set of resulting ensembles was calculated using validation data acquired from simulations employing the ESN architecture described in Section 5.3.

In summary, given minimal noise conditions, the largest data set comprising Zernike coefficients acquired using 8 source beacons was monotonic and returned the lowest AMSE for small field angles. However, architectural constraints may limit the performance of the

ESN. These, and other related issues, are discussed in Section 5.4.

7.3 Small field angles

Given the capabilities of the ESN method to predict spatiotemporal phase aberrations over a wide field, enhanced isoplanatic prediction was anticipated. The same configuration discussed in the preceding section was employed, i.e., minimal noise constraints, however angular separations were reduced to range from 0 to 25 μrad. The data ensembles from these propagation simulations were then used as input to an ESN; the results are presented in this section.

Simulations using 6 source beacons for small field angles were conducted. The results of using the ESN method on these data ensembles is given in Table 7.2.

Table 7.2 AMSE results for small field angles, 0 to 25 μrad.

Field angle, θ_{sep} (μrad)	0	5	10	15	20	25
AMSE	5.4 E-4	2.2 E-2	5.9 E-2	8.8 E-2	0.137	0.174

Table 7.2 shows a similar response to the analysis conducted over large field angles, where the AMSE is further reduced with a narrowing of field angle. Interestingly, but not unexpectedly, the monotonicity of the response increased with small field angles. Additional, narrow field simulations with 3 and 8 source beacon configurations were conducted, however similar responses to those shown in Table 7.2 were found.

7.4 Prediction by Fried's parameter and field angle

The Fried parameter was varied to evaluate the performance of the ESN method using different phase screen configurations, each generated using a different coherence length, r_0. The Fried parameter is defined by Equation 4.2 and represents a convenient parameter to characterise turbulence strength with respect to aperture size.

As with other simulations described in this chapter, the effects of turbulence from several source beacons were used to predict turbulence aberrations from a single target, with separations that exceeded the isoplanatic angle. Minimal noise constraints were used for this

evaluation, however the length of wavefront propagation (the height of a turbulent layer) was varied. These resulting data ensembles provided insight into ESN performance, in terms of propagation length.

Random ensembles from propagation simulations were used to train an ESN and the results from the ensuing predictions are shown in Figure 7.3, where two datasets are given. Each dataset is analysed separately in the following discussion.

Firstly, the set comprising Subfigures (a), (c), and (e) are the results from wavefront propagation at a simulated altitude of 12 km. As described in Subsection 6.2.2, wavefront propagation was limited to this altitude to meet spatial sampling requirements. Each set of results used phase screens generated with Fried's coherence lengths of, 20 cm, 10 cm, and 5 cm. Using a 1-m aperture, the resulting D/r_0 values are, 5, 10, and 20, respectively.

The set of source beacons used for ESN training and testing were maintained to serve as a basis for evaluation. As reported in the previous section, an 8-source beacon configuration provided the lowest AMSE for small field angles, given a constant Fried's coherence length, $D/r_0 = 10$. Similar results were recorded in this evaluation, however as r_0 was increased (less severe turbulence), and decreased (more severe turbulence), prediction performance improved and worsened, respectively. For example, the best result in this dataset is shown in Figure 7.3 (a), where 3 source beacons are used with minimal noise constraints, propagated over 12 km and over a field angle of 24 μ rad, and where $D/r_0 = 5$.

Secondly, Subfigures (b), (d), and (f), represent wavefront propagation simulated from an altitude of 8 km; all other parameters from the previous dataset are identical. In terms of monotonicity, Subfigure (b) is monotonically increasing with field angle. However for smaller values of r_0 (stronger turbulence), such as in (d) and (f), the majority of responses show monotonic behaviour.

The results from this section are consolidated in Figure 7.4, where a statistical average of AMSE predictions from 3, 6, and 8 source beacons are represented in terms of a single, r_0 response. Thus, the performance of the ESN method is proportional to Fried's coherence length, r_0, and to a lesser extent, the length of wavefront propagation.

For all simulations, the size of the input vector was reduced to prevent saturation of the ESN. Since the optimal number of input terms was two, the size of the maximum input vector, given by Equation 6.25, was 24. This resulted in additional processing, i.e., each group of two Zernike coefficients required individual presentation to the ESN, however an

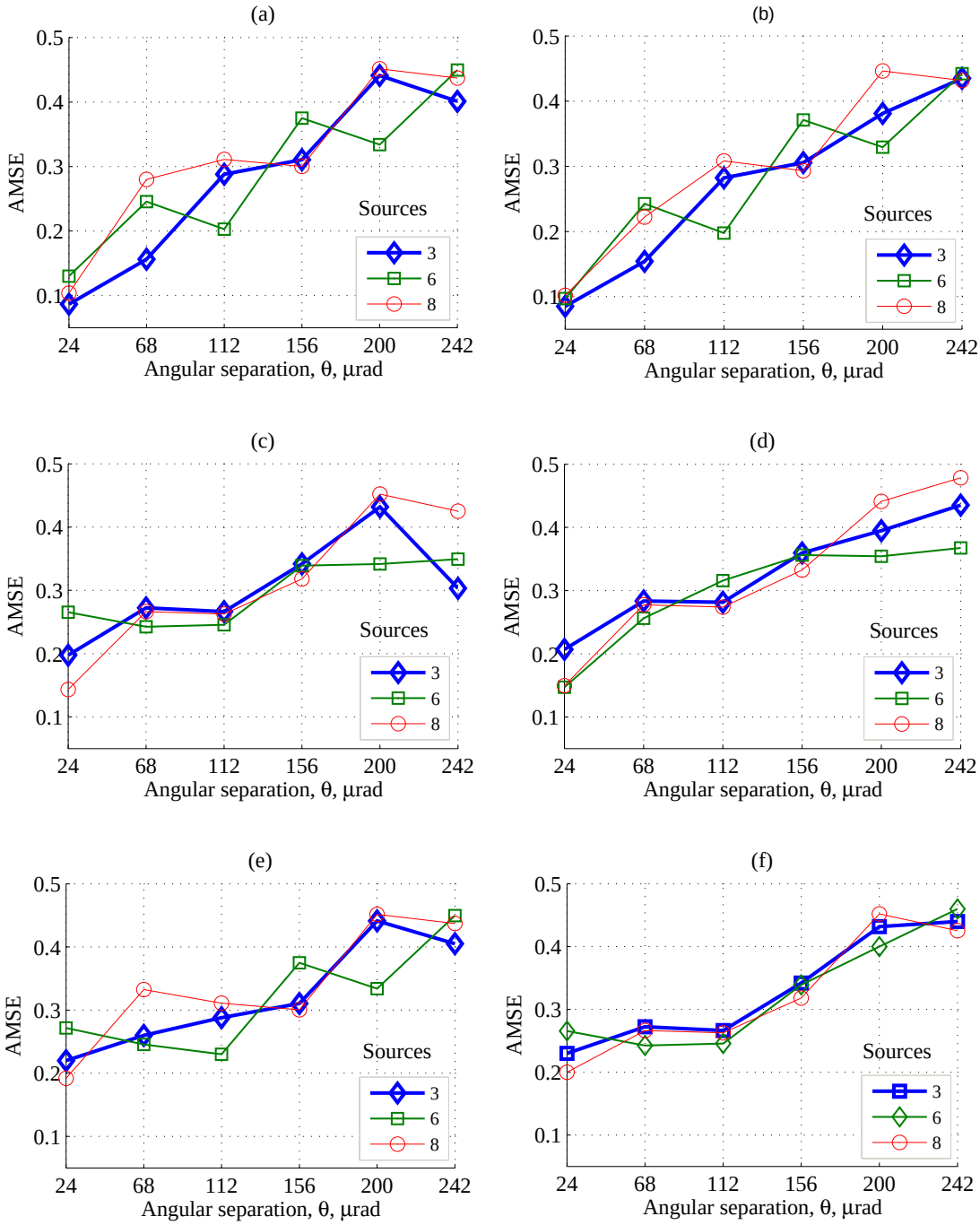

Figure 7.3 Wavefront averaged mean squared prediction error by angular separation, θ; (a) & (b) Fried's parameter = 20 cms; (c) & (d) Fried's parameter = 10 cms; (e) & (f) Fried's parameter = 5 cms. All simulation plots were generated using 1000 time samples, 136×10^3 photons, 80dB Gaussian noise, and separation, θ, of target and equally spaced sources, ranging from 24 to 242 μ radians; figures on the left are based on simulated wavefront propagation at a height of 12 km, the figures on the right are based on a height of 8 km.

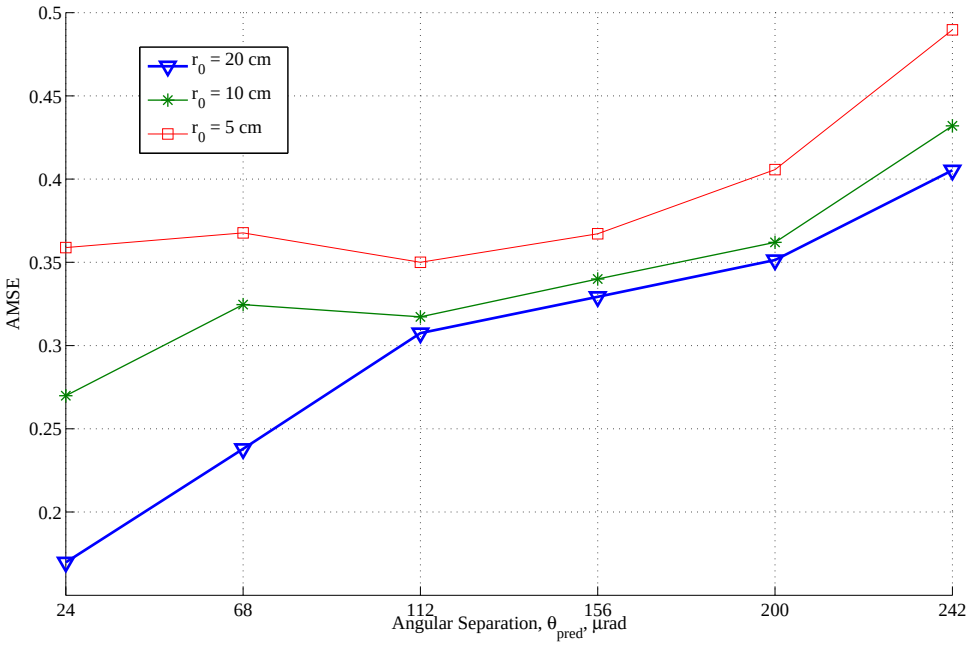

Figure 7.4 AMSE prediction error for a series of turbulence (Fried) parameters, r_0, over various field angles.

alternative was to increase the size of the ESN reservoir.

7.5 Performance evaluation with CCD noise constraints

Image sensor arrays, such as CCDs, are adversely affected by read noise. Read noise is an additive quantity and can be defined as a photon count error per pixel. When imaging faint, point-source beacons using a curvature-based WFS, images are subjected to read noise that result in Zernike coefficient errors. As noise variance increases, image quality is reduced, resulting in unreliable data. Read noise is modelled using Gaussian noise statistics with zero mean [26], and is defined by Equation 2.19.

The importance of including read noise in this evaluation is justified by the following analysis. The signal-to-noise ratio for CCD devices used to image faint astronomical objects can be characterised by [95],

$$\mathrm{SNR}_{\mathrm{CCD}} = \frac{N_p \eta}{\sqrt{N_p \eta + n_{pix}(N_s + N_D + N_R^2)}}, \tag{7.1}$$

where N_p is the photon flux, measured in photons per pixel per time interval t, η is the quantum efficiency of the CCD, n_{pix} is the number of pixels used for the SNR calculation, N_D is the dark current in electrons per n pixels per second, N_s is the background sky irradiance, measured in photons per n pixels, N_R is the read noise measured in electrons.

According to Howell [53], Equation 7.1 is typically used for *bright* source objects, i.e., when CCD gain and digitisation errors, σ_f, are regarded as negligible. In such cases, Poisson statistics dominate, and Equation 7.1 is reduced to,

$$\begin{aligned} \mathrm{SNR}_{\mathrm{CCD}} &= \frac{N_p\eta}{\sqrt{N_p\eta}} \\ &= \sqrt{N_p\eta}. \end{aligned} \tag{7.2}$$

However, when imaging *faint* source objects, gain and digitisation noise cannot be ignored. To allow for these conditions, Howell provides an alternative expression [53], given as

$$\mathrm{SNR}_{\mathrm{CCD}} = \frac{N_p\eta}{\sqrt{N_p\eta + n_{pix}(1+\frac{n_{pix}}{n_B})(N_s+N_D+N_R^2+G^2\sigma_f^2)}}, \tag{7.3}$$

where the term $(1+\frac{n_{pix}}{n_B})$ is an estimate of the background noise in the CCD, n_B is the number of pixels used to estimate mean background level, G is CCD gain, and the term σ_f^2 is the one standard deviation noise estimate introduced in the analogue to digital converter (ADC).

Careful application of Equation 7.3 reveals that imaging a bright object, given excellent seeing resulting in high image resolution, $(\frac{n_{pix}}{n_B} \leq 1)$, and with low background (sky) radiation, SNR figures in excess of 50 dB can be achieved [53]. However, given high gain, poor seeing conditions, and a high level of background radiation, the SNR results for a high-speed CCD[1], can be as low as 20 dB.

These examples helped justify the following performance analysis using noise affected CCD images. A normalised Gaussian distribution was used to simulate read noise conditions, and the adverse affects on CCD images employing the curvature WFS method, discussed in Section 4.4, were assessed.

Additional CCD noise, referred to in the literature as *dark current*, and generated by the

[1] Kodak CCD image sensor, KAI-0340DM

thermal generation of electrons, was considered. Electronic cooling assemblies, referred to as Peltier devices, are typically employed on CCDs. Such devices can effectively reduce dark current by up to 2 orders of magnitude, and residual dark current noise can be removed using *dark frames* [53]. Given this remedial procedure to minimise dark current noise [11], this topic is not addressed here.

7.5.1 Test procedure

The ESN was firstly trained with Zernike coefficient data ensembles acquired from images processed using the geometric wavefront sensor, with 80 dB additive read noise. Similarly, actual test data, in terms of Zernike coefficients for target objects, were acquired for comparison. Secondly, sources, comprising N image pairs, were acquired using the wavefront sensor discussed in Subsection 4.4.4.

Modelling of the CCD characteristics, i.e. SNR_{CCD} in Equation 7.3, was not performed. However, an analysis was conducted by inducing wavefront sensor images with Gaussian distributed read noise. Each pair of simulated, WFS images was distorted with a noise variance of σ_s. The variance of each undistorted image, σ_n, was used to calculate the SNR on each pair of WFS images. Individual simulations were conducted on each of seven resulting sets of noise induced WFS images, where the SNR of each set ranged between 80 to 20 dB.

Several sets of Zernike coefficients, each representing the statistical eigenfunctions of optical distortions to the optical path of a source beacon, were used to predict perturbations to the optical path of a target object. Gaussian noise profiles were generated and added independently to each WFS image pair. This comparative analysis was based on the angular separation of 3 to 8 source references to a single target, each with an equal separation of 112 μrad. This separation angle represented a *mid-point* between a range of angular separations, θ, listed in Table 6.1. The results were expected to provide insight into the ability of the ESN to discriminate noise induced WFS signals, for estimation of target wavefront distortion. Such conditions are common, for example, when viewing faint astronomical objects used as source beacons, requiring high CCD gains.

7.5.2 Results

In this evaluation, the wavefront prediction error is represented as a summation of the first eleven Zernike coefficients, with the piston term removed. Justification of this is given by

Hardy [49], who states that Zernike modes above Z_8 become increasingly more difficult to measure. The amplitude of successive Zernike coefficients is attenuated as a function of mode number. This is due to the $\kappa^{-8/3}$ shape of the modal spectrum [49], where κ is the power spectral density of the index of refraction fluctuations [117]. In consideration of this, Zernike modes two to eleven, i.e., $\{Z_2 \cdots Z_{11}\}$, were used in this assessment. The results, in terms of normalised, averaged MSE, are shown in Figure 7.5.

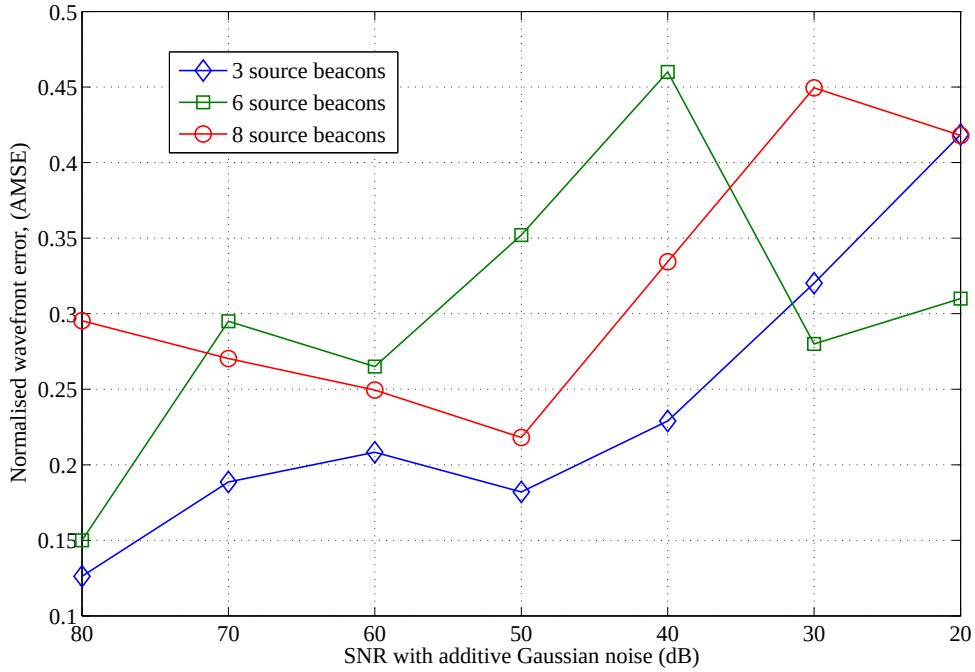

Figure 7.5 Normalised wavefront error for 3, 6, and 8 source beacons with each wavefront sensor image distorted by various levels of additive Gaussian noise.

Two observations can be made concerning these results. Firstly, a comparison of each response in Figure 7.5 would suggest fewer sources provide higher performance, and this appears consistent throughout the entire noise range. Considering that data from each additional source beacon adds 12 random variables to the ESN input, it is conceivable that additional inputs undermine the advantage of subsequent data points, especially considering the adverse effects of additive, Gaussian noise. Secondly, it is clear from this evaluation that wavefront error is, overall, adversely affected by Gaussian noise. The slight reduction in MSE at 50 dB for 3 and 8 beacon responses, and at 60 dB for 6 beacons, is thought to be due to improved generalisation, where additive noise to the inputs during training results in a smoothing, or regularisation [119], resulting in an overall improvement in performance. However, it is also apparent that excessive noise, i.e. SNR of less than 60 dB, has

a detrimental effect on three responses. While the 3 source beacon response is not strictly monotonic, it provides the highest order and overall best response (lowest MSE) of this ensemble.

Based on these results, the following comments can made. Firstly, evidence suggests the network is more tolerant to the effects of read noise if fewer source beacons are used. Since K additional beacons, each supporting M Zernike coefficients, results in $K \times (M+1)$ inputs, network overloading through saturation is thought to constrain performance. Increasing the size of the ESN reservoir can compensate for such conditions, however knowledge *a priori*, of this limitation is required. Further discussion on this topic is given in Subsection 7.7.1. Secondly, these results support what is intuitively known - small amounts of additive noise can improve network performance. However, when noise values result in $\text{SNR}_{\text{CCD}} < 50$ dB, network performance is significantly compromised. In practical terms, using the field configuration outlined in Appendix B, this SNR threshold equates to a moderate CCD gain setting,[2] suitable for imaging relatively bright source beacons, for which $m_v \leq 5$.

7.6 Performance evaluation with photon noise constraints

The probability of imaging bright NGSs, over a single FOV, is proportional to the percentage of sky coverage [90]. Wide-field imaging is used in this study, however other factors, such as acquisition of sufficient photons to accurately determine wavefront aberrations from several source beacons, each possibly of different magnitudes, are considered in this section. For example, sampling rates that exceed the Greenwood frequency, f_G, are used to minimise atmospheric decorrelation. While these rates are achievable using commercial CCDs, e.g., sample periods ≤ 5 ms, photon flux, as shown by Equation 2.24, is constrained. Photon starvation results in *photon noise*, and degrades WFS performance [19]. Of specific concern are NGSs of magnitudes, $m_v \geq 6$. These considerations are addressed in this section, where ESN performance for wavefront prediction is evaluated for various photon starved conditions.

In terms of the WFS used in this study [19], photon noise can be modelled by normalising both intra-focal and extra-focal images and specifying a photon count. The photon count is a combination of the photo-diode sensitivity and the integration period used in the exposure. The Poisson distribution is used as a basis for the photon noise model, and the effects of random photon events, given a total number of photons collected per image over a sample

[2] Kodak CCD image sensor, KAI-0340DM

period, Δt, is simulated. To demonstrate the effectiveness of this method, a planar wavefront was perturbed using a phase screen, and four images were produced using a range of photon events. A collection of these four images is shown in Figure 7.6, where the inverse of each sub-image is displayed for clarity.

In Subsection 2.2.5, it was shown that photon flux is proportional to the photon count captured by a detector, such as a photo-diode, over time. However, the arrival of photons is a random event that can be described in terms of probability using the Poisson distribution. In the case of the curvature WFS, the photon noise was simulated using a Poisson noise model, where the conditional probability of the arrival of photons, $P(\cdot)$, is given as [19]

$$P(I(x,y)|i(x,y)) = \prod_{\forall(x,y)} \frac{\exp[-i(x,y)]i(x,y)^{I(x,y)}}{I(x,y)!}, \tag{7.4}$$

where $i(x,y)$ and $I(x,y)$ are intensity measurements before and after the addition of photon noise, respectively.

To represent the effects of photon noise on the curvature WFS, Equation 7.4 can be applied independently to both the intra-focal WFS image, $i_-(x,y)$, and the extra-focal WFS image, $i_+(x,y)$, as described mathematically by Yong [19]

$$I_+(x,y) = i_+(x,y) + n_+(x,y), \tag{7.5}$$

and

$$I_-(x,y) = i_-(x,y) + n_-(x,y), \tag{7.6}$$

where $i_\pm(x,y)$ and $I_\pm(x,y))$, are intensities before and after the addition of photon noise, $n_\pm(x,y)$, respectively.

Application of Equations 7.5 and 7.6, in terms of simulating inducing photon noise as part of propagation model, provided the basis of a quantitate analysis on the performance of an ESN to estimate target perturbations. Since irradiance measurement planes, $I_+(\cdot)$ and $I_-(\cdot)$ are also for the geometric WFS, simulations with induced photon noise was appropriate for both WFSs.

[2]Kodak CCD image sensor, KAI-0340DM

7.6.1 Test procedure

The ESN predictor was tested using random data ensembles, comprising photon noise induced Zernike coefficients generated by the method described in the preceding section.

Photon noise was incrementally applied to WFS images over the range listed in Table 7.1. Exponential increments were used, as is common for such analyses, and provided a wide range of intensity distributions for simulations. Several examples of the simulated effects of photon noise are shown in Figure 7.6.

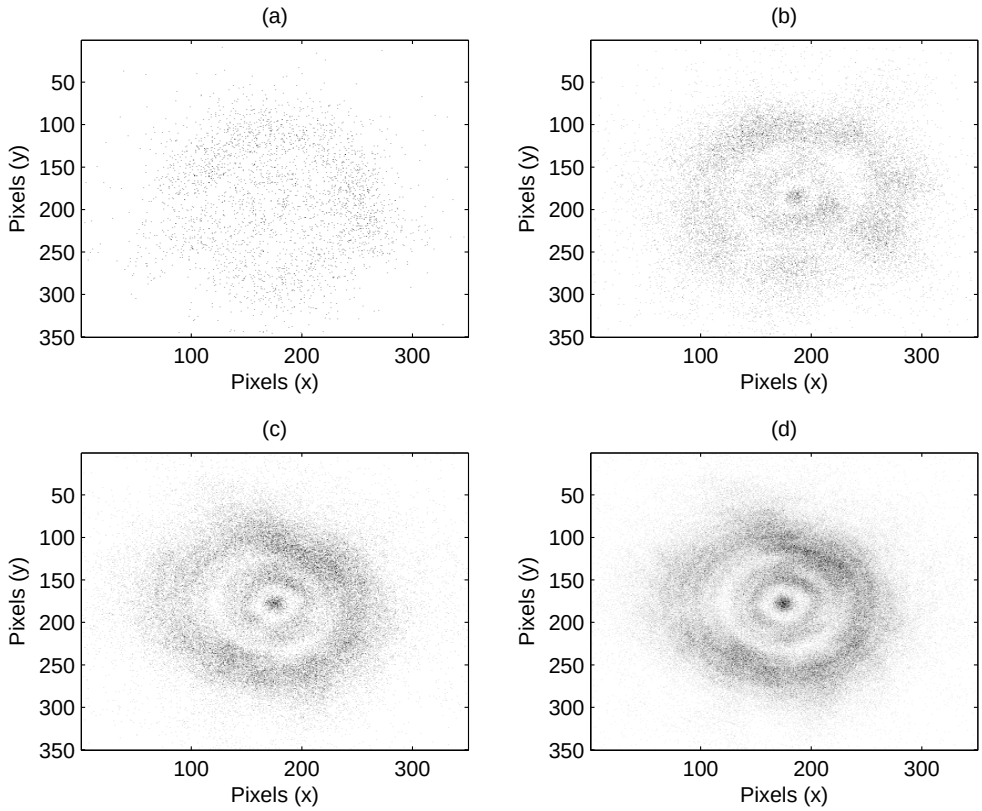

Figure 7.6 Effects of imaging a point-source object with photon noise: (a) 2.91×10^3 photon events, (b) 2.2×10^4 photon events, (c) 5.99×10^4 photon events, (d) 1.63×10^5 photon events.

Additionally, Zernike coefficients were individually trained and tested, rather than re-scale and accumulate each mode for wavefront phase reconstruction, as was performed in Section 7.5. The objective here was to assess the performance of the network on individual Zernike coefficients. Given that higher order modes are arguably subjected to higher order dynamics due to lower inertia [134], it was theorised that these modes would be more adversely

affected by noise, compared to lower order terms, such as tilt, $\{Z_2, Z_3\}$. Therefore, higher-order modes should, according to this hypothesis, be more difficult to predict.

7.6.2 Results

The spatiotemporal training method described in Subsection 6.3.2 required a minimum of three source beacons to estimate Zernike coefficients in anisoplanatic regions. However, the performance of the ESN was also tested using six and eight source beacons. The rationale was that an increase in the number of source beacons was expected to provide more data points and improve prediction. The results of simulations for each source configuration are shown in Figure 7.7.

Subfigures shown in Figure 7.7 are arranged in horizontal pairs, and each pair results from a different number of sources. Identical datasets were used for each subfigure pair, however the left plots ((a), (c), and (e)), highlight individual Zernike coefficient responses. The legend shown in Figure 7.7 applies to the first 11 Zernike coefficients (ignoring piston, a_1), associated with left plots. Lastly, symbols have been used for the first four coefficients, $a_2, a_3, \cdots, a_5$, to highlight low-order modes.

Simulations performed with high photon noise, e.g., 150 photons, resulted in poor performance; this was expected due to the inability of the WFS to return a repeatable wavefront map over a series of short exposures, given identical phase perturbations. As photon noise was decreased by increasing photon flux, improvement in network performance was noted; this is reflected in subfigures (a), (c) and (e) in Figure 7.7. While the predicted error for each coefficient varied in response to increments of photon flux, a monotonically decreasing result is evident. Successive increases in photon flux (not shown in Figure 7.7) further improved network performance, where the response for 8 source beacons converged to a minimum AMSE≈ 0.1 for 8.32×10^3 photons.

In summary, phase perturbations were measured simultaneously using the geometric WFS, where defocused images from several source beacons were distorted by photon noise. Each random data ensemble were used as input to an ESN for prediction of target aberrations. The results presented in this section can be summarised as follows:

- The geometric wavefront sensor achieves lower wavefront estimation errors compared to the curvature sensor, especially for wavefront propagation over short distances [19]. Initial tests confirmed this, and the geometric WFS was used in subse-

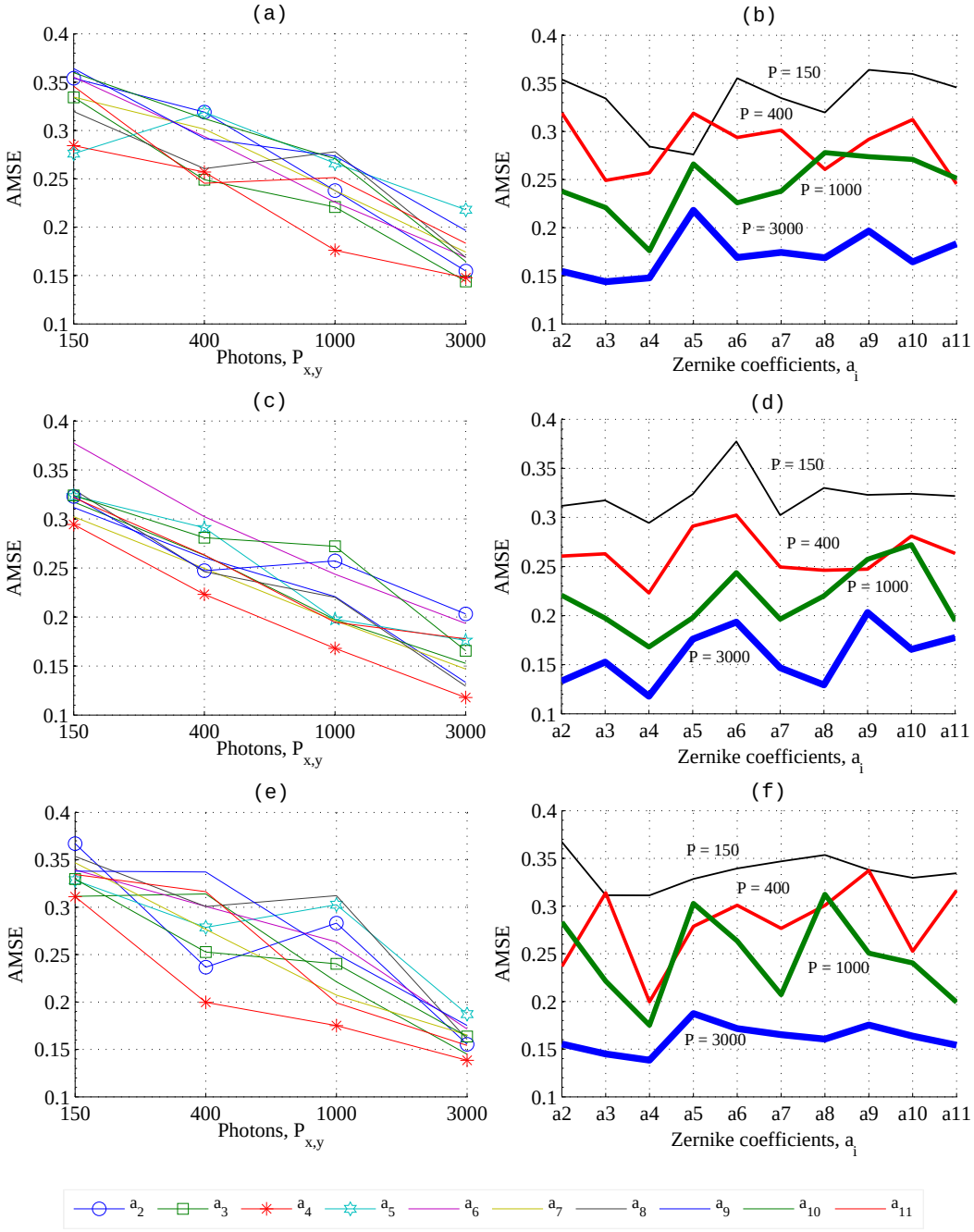

Figure 7.7 AMSE prediction of Zernike coefficients using 3, 6, and 8 source beacons, at an angular separation of 240 μrad, with approximately 150, 400, 1000 and 3000 photons: (a) Zernike coefficients for 3 source beacons; (b) Photon noise profiles for 3 source beacons; (c) Zernike coefficients for 6 source beacons; (d) Photon noise profiles for 6 source beacons; (e) Zernike coefficients for 8 source beacons; (f) Photon noise profiles for 8 source beacons.

quent simulations.

- Poisson noise severely affects WFS performance, and this was incorporated in propagation simulations for ESN evaluation to predict target aberrations. Performance of the ESN confirmed a monotonically decreasing prediction error, as photon noise was reduced.

- Prediction errors for individual Zernike modes assessed in this study were similar, i.e., using the ESN prediction method, no distinctions between low- or high-order coefficients were noted.

- A marginal improvement in the rate of convergence to a minimal AMSE was noted in the case for 8 source beacons, when minimal photon noise was applied.

Images of faint astronomical objects, distorted by photon noise, require high gain sensors for wavefront sensing. Such conditions, however, can result in a combination of noise sources, i.e., photon and read noise, where the latter is generally applicable to CCD image sensors. As discussed in the preceding sections, such individual noise sources degrade prediction performance. However, since a combination of noise sources is often encountered in field observations, this topic is discussed in the following section.

7.7 Network performance with combined noise effects

In this section the results of simulations employing a combination of both photon and CCD read noise are presented. The results, in terms of normalised, averaged mean squared error (AMSE), represent predictions of a series of Zernike coefficients, $a_2, a_3, \cdots, a_{11}$, using image data from three source beacons, where each image pair presented to the WFS is distorted by a combination of photon and CCD read noise conditions.

The Poisson distribution is used to simulate low-light conditions, whereas the Gaussian distribution is used to represent CCD read noise; both distributions are defined in Subsection 2.2.5. The effect of combining both noise sources provides insight into the reliability of the ESN prediction method detailed in Section 6.3, and its use with empirical data. For example, phase data generated either in the laboratory or from field observations using commercial, high frame-rate CCD cameras, can be compared with simulation data to verify model parameters. Four examples of CCD images, distorted by various combinations of photon starvation and read noise, are shown in Figure 7.8.

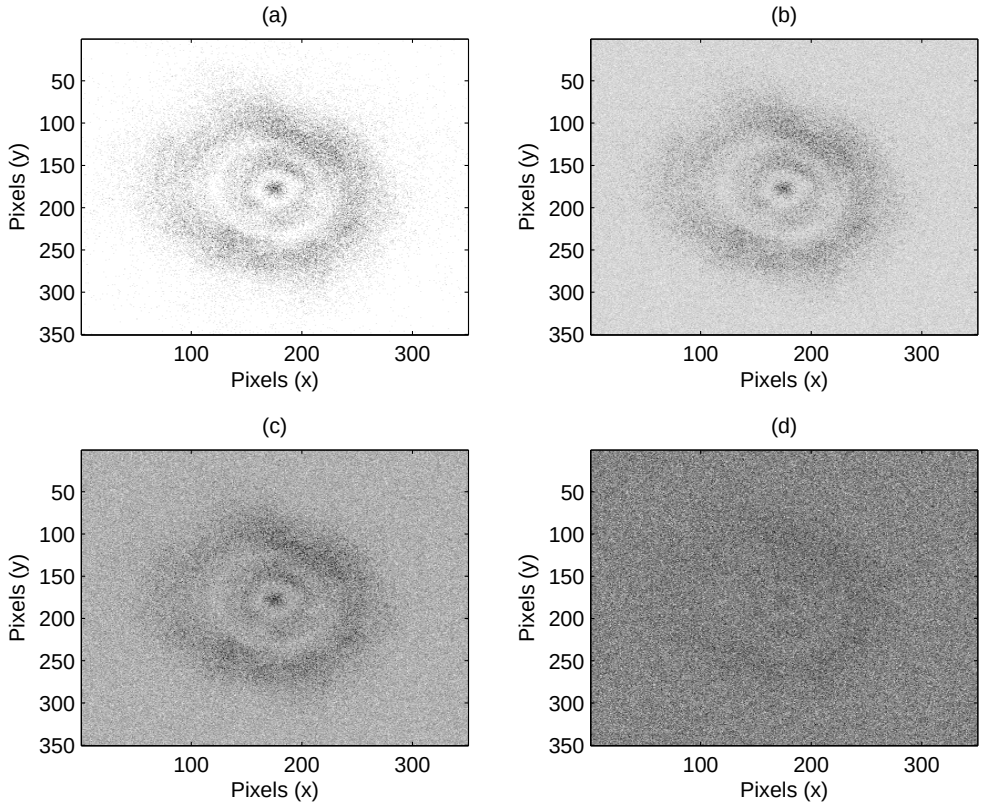

Figure 7.8 Series of negative images showing the effects of combined Poisson noise, 6×10^4 photons, and read noise, for SNR values of: (a) 40 dB, (b) 11 dB, (c) 0 dB, and (d) -28 dB.

Each image in Figure 7.8 was generated using 6×10^4 photon events, and inverted to highlight combined noise characteristics. Moderate to extreme levels of Gaussian noise ($40 \leq \text{SNR}_{\text{CCD}} \leq -28\,\text{dB}$) have been added to illustrate harsh environmental conditions that can be simulated for wavefront acquisition.

The test procedure performed in Subsection 7.6.1 was repeated, however simulations combined both read and photon noise, and the analysis was consolidated using only three source beacons. More specifically, each propagation simulation used incremental photon and read noise values from a parametric array. The resulting random ensembles were used to train the network. A separate set of random ensembles, employing the same noise characteristics, were used to test ESN performance. These results comprised a matrix of target predictions, in terms of the normalised, AMSE metric. The photon count ranged between 3000 (moderately high noise) to 2.2×10^4 (moderately low noise), and additive CCD read noise were combined from 40 dB (moderately high noise) to SNR of 80 dB (low read noise), in 20 dB

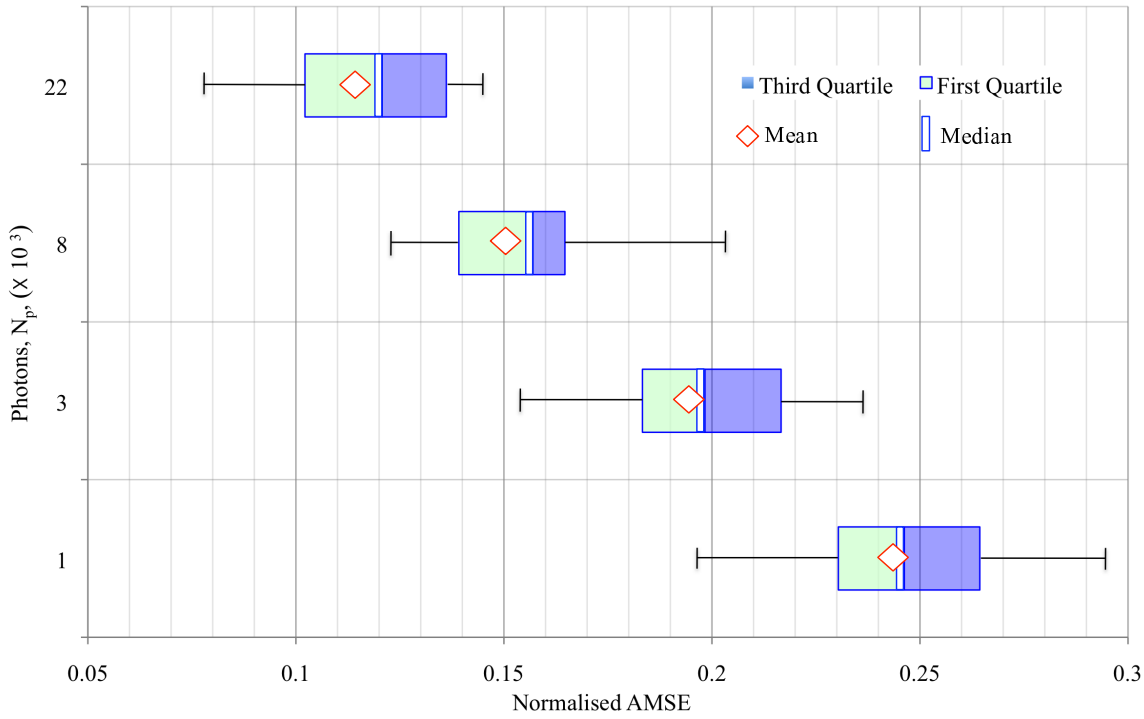

Figure 7.9 Prediction error of Zernike coefficients, $\{a_2, \cdots, a_{11}\}$, employing three source beacons. The normalised AMSE was calculated, based on a series of SNR values: 80, 60, and 40 dB over the first 10 Zernike modes, where Photon noise was varied using approximately 2.2×10^4, 8×10^3, 3×10^3, and 10^3 photon events. Each series of tests showed similar results.

increments.

Figure 7.9 summarises ESN performance, where read noise is incrementally increased over the range, 80 to 40 dB, and photon noise is applied to WFS images, resulting in noise-induced Zernike coefficients. Commensurate with the previous study in Section 7.6, where only photon noise was applied, this study showed that ESN prediction is dominated by photon noise, given the SNR range used in this evaluation.

Intuitively, it is reasonable to assume that severely distorted wavefront images will be uncorrelated between each time-step. The rate of decorrelation can be related to the ESNs inability to learn a sequence of temporal wavefront phase signals, in terms of Zernike coefficients. Training on a combination of read and severe photon noise results in random sequences, where any underlying pattern representing the temporal evolution of a series of wavefront aberrations, is lost. As photon flux is increased, the ESN regains the ability to respond to the dynamic spatiotemporal sequences the network was trained for, allowing the formulation of an output response that has a degree of similarity to distortion affecting

a target. Thus, this study has shown that ESN performance is susceptible to both photon and read noise, however photon noise is dominant over the ranges used in this evaluation. However, as shown in Figure 7.5, excessive read noise, resulting in low SNR values, i.e., < 40 dB, is also detrimental to performance.

7.7.1 Input data dependency

The following analysis on input dependencies extends the discussion on ESN optimisation presented in Section 5.4. In addition to network optimisation using specific design parameters, the correlation of time series data input to the network is considered in this subsection.

Refinements to ESN performance can be made by tuning network parameters, such as the density of the dynamic reservoir (DR). However, in this subsection consideration is given to training the network with correlated data sequences. Network performance was assessed in two ways.

Firstly, in Section 7.4, ESN performance was shown to be proportional to the number of data streams (source beacons). However, if the number of ESN source inputs exceeded 8, ESN performance diminished. Experimentation suggested that the processing of input data into loosely coupled systems was restricted due to insufficient neurons. Increasing the number of neurons is a straightforward solution, however as shown in Figure 5.4, simply increasing the size of the DR can have an adverse effect on performance.

Secondly, since the ESN is trained using time series data composed of the Zernike orthonormal basis set, information encoding is maximised in terms of independent, modal expansions. This expansion set is defined in Subsection 3.3.1, and the first 20 modes of the Zernike series were acquired; approximately half of these modes were used to reconstruct the SVPSF. Maximising relevant information for this process is addressed in the information-theoretical method known as *information bottleneck* [130]. However, since the Zernike series is sub-optimal, as shown by off-diagonal entries in the Zernike covariance matrix in Table 4.3, the feasibility of further reducing prediction error by employing mutual information between pairs of Zernike coefficients was investigated. The method used to determine this is outlined in the following procedure.

An ESN can be configured to predict either a single Zernike coefficient, or multiple coefficients. For example, Zernike mode a_2 can be predicted for a target using the same mode perturbations from multiple sources, $\mathbf{a}_2$, and angular separations to the target, $\boldsymbol{\theta}_{\text{Sep}}$. Addi-

Table 7.3 Zernike coefficient dependency analysis for $K = 3$, $D/r_0 = 10$, and $\boldsymbol{\theta}_{\text{Sep}} = 24\,\mu$rad.

Set	K-**sources** $(\mathbf{a}_i = K \cdot a_i)$	**Target prediction**, MSE, $\bar{b}_j$					
		$\bar{b}_2$	$\bar{b}_3$	$\bar{b}_4$	$\bar{b}_5$	$\bar{b}_6$	$\bar{b}_7$
1	$\{\mathbf{a}_2\}$	0.0617	-	-	-	-	-
2	$\{\mathbf{a}_2, \mathbf{a}_3\}$	0.0484	0.0802	-	-	-	-
3	$\{\mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$	0.0867	0.0879	0.1052	-	-	-
4	$\{\mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4, \mathbf{a}_5\}$	0.0512	0.0799	0.0929	0.1215	-	-
5	$\{\mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4, \mathbf{a}_5, \mathbf{a}_6\}$	0.0527	0.0818	0.0805	0.1457	0.1458	-
6	$\{\mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4, \mathbf{a}_5, \mathbf{a}_6, \mathbf{a}_7\}$	0.0832	0.0916	0.1058	0.1226	0.0958	0.1491

tional modes can be included to contribute to such predictions for a_2, by including N input modes from the same sources, $\mathbf{a}_i$, where $i = \{2, 3, \cdots, N+1\}$. It was theorised that in the absence of mutual coupling, identical results should be returned for both single-mode prediction, and multiple-mode predictions. Conversely, small, but measurable *leakage* between specific modes, e.g., $\{a_2, a_3\}$, and $\{a_7, a_8\}$, should adversely affect ESN performance, commensurate with the degree of mutual coupling between modes.

The modal predictor shown in the system model of Figure 6.1 was modularised to predict M output coefficients, from N input coefficients. As outlined in Section 6.1, k sources were required, where $k \in \{3, 4, \cdots, K\}$, over θ_k field angles, for spatiotemporal prediction of a target Zernike coefficient, b_j, where the letter representing the coefficient is changed to differentiate an input coefficient from an output. Table 7.3 lists a succession of inputs comprising sets of Zernike coefficients. For example, Set-1 employs a set of the Zernike *tilt* coefficient, $\mathbf{a}_2$, where $\mathbf{a}_2 = k \cdot a_2$. This set is used to predict a single, target coefficient, b_2. The averaged result, MSE $\hat{b}_j$, is shown in column 3. For Set-2, two consecutive Zernike modes, $\mathbf{a}_2$ and $\mathbf{a}_3$, are used to predict b_2 in column 3, and b_3 in column 4. In theory, the coupling between coefficients should be minimal, resulting in a consistent MSE, irrespective of the number or combination of Zernike coefficients input to the ESN.

In Table 7.3, the MSE on the $\hat{b}_j$ diagonal is increasing with Zernike mode, j. This suggests that higher modes are more difficult to predict than lower modes. However, there appears to be some inconsistency with the results show in the columns for $\hat{b}_j$; e.g., inclusion of lower

modes returned mixed results for higher modal predictions. To assess these results in terms of residual wavefront error, the procedure was repeated, however individual AMSE results were scaled and accumulated to calculate AMSE, $\varepsilon_{\mathrm{R}}^2(\cdot)$ over N Zernike modes.

Consider a time series ensemble of Zernike coefficients, $\mathbf{a}_i$, where $i \in \{2,3,\cdots,11\}$ and each set is assigned to three source beacons, i.e., $K = 3$. The resulting source ensemble is input to an optimised ESN for prediction of target aberrations $\mathbf{b}_j$, where $j \in \{2,3,\cdots,11\}$. Combinations of i are used to find the minimum, residual wavefront error, AMSE, $\varepsilon_{\mathrm{R}}^2(D/r_0)$, where $D/r_0 = 10$. An additional condition, shown in Table 7.3, is that the specific mode being predicted is represented by its associated input mode. For example, if the network is required to predict Zernike coefficient b_6, coefficient a_6 from K source beacons must be used to both train and test the network.

A series of simulations were conducted, where the number of Zernike mode inputs to the ESN were increased independently to the number of outputs. The residual wavefront error, $\varepsilon_{\mathrm{R}}^2(10)$, in terms of normalised AMSE, is shown in Figure 7.10.

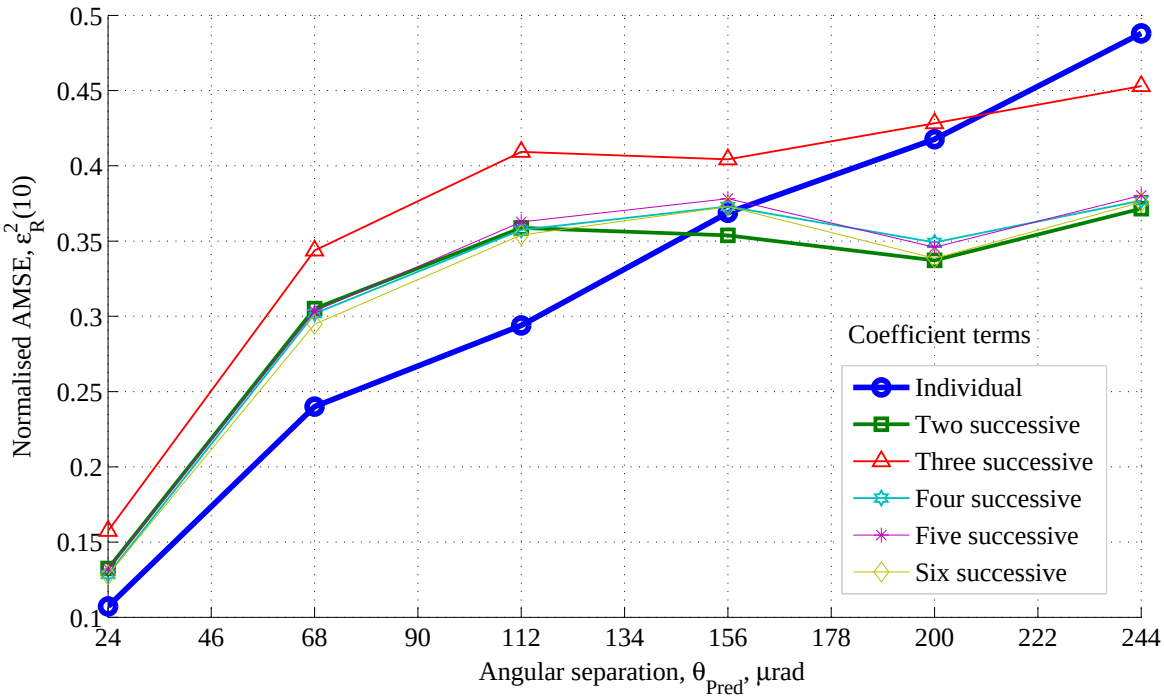

Figure 7.10 Wavefront performance analysis using individual and successive orders ($N = \{1,2,\cdots,11\}$), contiguous Zernike coefficient input terms.

Each set of results in Figure 7.10 shows a similar response to that presented in Section 7.4, i.e., prediction performance in terms of MSE is dependent on angular separation, $\text{MSE}_{b_j} = f_{b_j}(\theta_{\text{Pred}})$. However, by defining network input as a function of Zernike modes, where DR size was proportioned accordingly, significant changes to network performance were noted.

Furthermore, structuring the ESN with individual Zernike modes resulted in the best overall performance. This is shown in Figure 7.10 where the lowest AMSE results were obtained for $\theta_{\text{Pred}} \leq 150\,\mu$rad, and the overall response was monotonic. However, combined Zernike modes showed mixed results.

Two additional observations can be made. Firstly, the Kolmogorov spectra defines more power in lower wave number components [116]; this is commensurate with increased strength in tilt Zernike modes and their dominance ($\approx 86\%$), in terms of wavefront error. Individual predictions of tilt terms, $\{a_2, a_3\}$, have shown the largest reduction in terms of AMSE. Secondly, since the Zernike modes are sub-optimal, the combined effects of Zernike modal predictions would appear to have a detrimental effect on wavefront error. An example of this is shown in the 4^{th} column of Table 7.3, where b_3 is predicted by Set-2 and also Set-6; the former includes an orthogonally independent input term, a_2, however the latter includes a correlated term, a_7. Thus, the highest prediction error of 0.0916 in the 4^{th} column suggests that the inclusion of a_7 is detrimental to the prediction of b_3.

This analysis has shown that even small correlations between Zernike coefficients can degrade network performance, and optimal ESN performance is achieved using a single input and output configuration. Furthermore, independent ESN predictors can be individually trained for specific modes and configured as a parallel computing structure to reduce computational complexity.

7.8 Performance summary

This chapter has detailed the performance of the ESN method using the AMSE metric. Observational and physical parameters, such as Fried's parameter length, r_0, angular separations, θ_{Sep}, and measurement noise, in terms of photon flux, N_p and CCD read noise, have been used to assess the performance of this method. These parameters and associated ranges used in simulations are summarised in Table 7.4.

The column labelled `AMSE` characterises the performance of each parameter in terms of average mean squared error. For example, the symbol $\downarrow$ indicates a reduction in AMSE,

Table 7.4 ESN performance summary for modal prediction.

Performance parameter	Symbol	Unit	AMSE	Range / Notes
Angular separation	θ_{sep}	μrad	$\uparrow$	$0 \rightarrow 244$
Fried's parameter	r_0	m	$\downarrow$	$0.05 \rightarrow 0.20$
Read noise (Gaussian)	SNR	dB	$\downarrow$	$20 \leftarrow 80$
Photon flux	N_p	J[a]	$\downarrow$	$150 \rightarrow 2.2 \times 10^{4}$[b]
Source beacons	K	-	$\downarrow$	Valid for $K \leq 8$
Zernike input combination	$\boldsymbol{a}_i$	-	$\uparrow$	$\forall i \in \{2,3,\cdots,M\}$

[a] One photon carries 3.97×10^{-19} Joules of energy when $\lambda = 500\,\text{nm}$
[b] Using 8 source beacons, the AMSE converged at 0.1 with 8.32×10^{4} photons

thereby enhancing ESN performance, whereas the symbol $\uparrow$ represents an increase in AMSE and subsequent degradation in ESN performance.

In summary, performance of the ESN method gradually diminishes over wide field angles and is limited by the effects of noise, specifically read and photon flux. However, while the former can be addressed through additional training, global network performance can be optimised using ESN design parameters.

Chapter 8

Conclusions and Future Research

The principal contribution detailed in this research is the ESN method used to improve anisoplanatic, astronomical imaging. Other contributions include a new wavefront sensor adaptation for acquiring simultaneous perturbations from multiple source and target objects, the development of a spatiotemporal algorithm for training a recurrent neural network, a method that interpolates individual Zernike modes for construction of the spatially variant PSF, and an analysis of Taylor's frozen turbulence hypothesis, incorporating a practical method to determine its limitations.

The first part of this chapter provides a summary of these contributions, and concludes the research presented in Chapters 4 to 7. The second part of this chapter presents a summary of future research, based on several topics that are closely related to this research.

8.1 Conclusions

A method for improving astronomical images of science objects, distorted by wavefront propagation through a turbulent atmosphere, is addressed in this dissertation. The primary focus has been to minimise the anisoplanatic PSF error over a wide spatial field. This was initiated by the use of wavefront data from multiple source beacons, and facilitated by the adaptation of a recurrent neural network. The proposed ESN method trains a specialised, recurrent neural network to predict spatiotemporal wavefront perturbations over wide angular separations. Since the research conducted is multidisciplinary, existing research in relevant areas such as adaptive optics, image restoration, and reservoir computing, lays the foundation for extensions in these areas. This material is presented in Chapters 1 to 3. In

developing the ESN method, research was completed in several stages.

Firstly, two curvature-based wavefront sensors were adapted to simultaneously acquire wavefront maps from multiple source beacons. By combining both zonal and modal operations in a single wavefront sensor, spatial measurements of source objects over the pupil facilitated the simultaneous measurement of phase perturbations from independent reference objects. These measurements provided spatial and temporal wavefront data for training, testing and verifying the ESN method. The adaptation of the curvature-based wavefront sensor, introduced in Chapter 4, was employed to acquire empirical wavefront data during field observations, and is integrated in a simulation framework presented in Chapter 6.

To assess the constraints of the simulation model, empirical data ensembles were acquired from several observation runs. These ensembles were used to test Taylor's frozen turbulence hypothesis to ascertain the extent of temporal decorrelation of the atmosphere. Essentially, the ESN would need to work harder to maintain accurate prediction of target wavefront maps if the metapupil were expanded. The relationship between the field-of-view, the height of a turbulent layer, and temporal decorrelation of the atmosphere, in terms of a set of modal expansions, was formulated. This relationship was used as a primary aid to establish a correlation between the temporal flow of "frozen" turbulence, and the spatial evolution of aberrations over the pupil. A time series analysis was conducted on Zernike coefficients using empirical data; this included tilt, defocus, and astigmatism wavefront aberrations. The results of this analysis concluded that chaotic behaviour was inherent in each mode, and that each time series was therefore deterministic. This result justified the use of a recurrent network for prediction of target aberrations.

Secondly, reservoir computing is introduced in Chapter 5 and was used as the primary structure for spatiotemporal prediction. Network design parameters were identified and network performance was analysed using an ESN. To facilitate a range of computationally intensive tests, a scaled version of the ESN predictor used in the final evaluation was employed for this analysis. The results from this analysis established guidelines for the design of an optimised ESN; improvements both in terms of network efficiency and performance were demonstrated. Noise induced data ensembles were used to both train and test the network; initial tests using noise induced wavefront data showed some performance improvement with low levels of CCD read noise. Network response to photon noise however, was monotonic across a wide range of noise levels. This topic is comprehensively addressed in the evaluation of the ESN method presented in Chapter 7.

Thirdly, a simulation framework was created to propagate multiple wavefronts through a dynamic, perturbing medium, resulting in time-varying data ensembles. These ensembles are used in conjunction with a spatiotemporal algorithm to train an ESN for prediction of target phase perturbations. A verification model is presented in conjunction with a detailed description of the simulation environment in Chapter 6.

Lastly, the results from an analysis using random ensembles were required to evaluate performance of the ESN method employing random ensembles. Such an analysis was conducted in Chapter 7. Two sources of noise commonly experienced in astronomical imaging were generated and applied to image data, thus emulating both moderate and extreme noise induced measurements of wavefront data. These results showed that by employing three or more source references over an angular separation of of 24 μrad from a target, and given mild turbulence with Fried coherence length of 20 cm, the normalised mean squared error of low-order Zernike modes could be estimated to within 0.086. Furthermore, the employment of additional references in excess of 8 source beacons did not show a significant performance improvement. However, prediction error was primarily dependent on field separation, since temporal decorrelation, constrained by the Taylor hypothesis, appeared to be a contributing factor. Additional network training partially addressed these issues.

Existing deconvolution algorithms used the spatially variant PSF to restore extended astronomical images. A block processing algorithm is employed to restore anisoplanatically distorted images in Appendix A. An enhancement to this algorithm was proposed to reduce boundary distortion over adjacent isoplanatic patches. This new method is outlined in Appendix A.

8.2 Future research

Suggestions for future research on this topic are presented in this section. These suggestions have arisen during the course of this research, however they have yet to be fully investigated. These include:

1. The algorithms developed for this work were not optimised. As a result, simulations, particularly for wavefront propagation, often took in excess of 120 hours to complete. In comparison, ESN training contributed to 2% of the total propagation simulation period. Benchmarking software from Mathworks MATLAB® was used to isolate routines contributing to long execution times. The results from this analysis showed that

wavefront propagation routines made extensive use of the 2D FFT function, whereas the prediction routines showed that the Moore-Penrose pseudo-inverse function was consuming the majority of computational bandwidth. Optimisation of these functions would result in significantly shorter simulation periods.

2. The simulation environment described in Chapter 6 was restricted to the memory capacity of the underlying computer hardware. For example, a PC supporting a 32-bit datapath and 8G-bytes of memory limited the size of a phase screen to 4096^2. However, a 64-bit architecture allowed 8192^2 phase structures to be implemented. As the phase screen was displaced with respect to a 'fixed' aperture, time series wavefront data was acquired. However, a limit existed as to the number of sample acquisitions conducted, w.r.t. to simulated wind speed; this was particularly restrictive over large field angles. For example, as angular separations increased beyond 200 μrad, sampling was limited to 500 iterations. Implementation of a dynamic phase screen, where the rate of continuous, turbulent flow could be regulated over large field angles, would address this limitation.

3. Additional training information could be employed by the ESN to further improve performance. For example, scintillation wavefront data could be used in conjunction with phase measurements to further minimise MSE prediction error of target wavefronts outlined in this study. Ragazzoni *et al.* [110] suggested these improvements by employing a single-image, curvature WFS. Furthermore, it may be possible for an ESN to learn to associate scintillation patterns with optical phase using combined data from multiple source beacons.

4. The Zernike series is an extremely useful basis set for describing aberrations over a circular aperture, however other basis functions exit and may be more suitable for prediction. An extension to this research could investigate the direct modal extraction of aberrations, in terms of either Zernike basis, the Karhunen-Loéve expansion, or an alternative basis set, using an ANN for classification. A single sensor configuration, similar to that adopted by Ragazzoni *et al.* [110], could be employed for this purpose. An immediate advantage of a single-image WFS, compared to geometric or curvature WFSs, is the maximal utilisation of light from faint source objects.

5. A hardware implementation of the ESN method is required for a practical application of deconvolution from wavefront sensing. Such an implementation would provide simultaneous wavefront data streams, in conjunction with distorted image data. Operating in an open-loop configuration, hardware deconvolution routines could be used

in conjunction with a trained ESN hardware module. Such an implementation is proposed for real-time image correction over a wide field-of-view.

6. A brief discussion on the performance of linear predictors, in comparison with ANNs, was given in Subsection 5.3.4. In addition, time-delayed feedforward networks were compared to the performance of ESNs in related work [142]. However, a more comprehensive comparison is required over a wider range of networks, including existing methods discussed in Section 4.6, for estimating the SVPSF.

7. High order Zernike coefficients are more affected by diffraction, compared to low order terms such as tilt. The simulation model used for multilayer wavefront propagation discussed in Section 6.2 employs a geometric model, where the effects of Fresnel diffraction between layers are ignored. As discussed by Hardy [49], a geometric propagation model is appropriate for weak turbulence models, however for mild or sever turbulence the effects of diffraction must be considered. The multilayer propagation simulations detailed in Section 6.2.4 should be repeated using Fresnel diffraction. The resulting data ensembles should be employed for predicting high-order Zernike terms and the results compared with the respective high-order terms detailed in this thesis.

A primary requirement however, for any or all the above mentioned research extensions, is access to empirical data. Much of the work completed in this research would not have been possible without analysis of data from either field or laboratory work. Thus, empirical data provides the basis for simulations that can be reinvested in methods, such as the ESN approach.

Appendix A

Restoration of Astronomical Images

In this appendix, two widely used image restoration methods are reviewed: Tikhonov regularisation and the expectation maximisation algorithm. These methods are applied to wide-field astronomical images, distorted by the effects of inhomogeneous random media. The distortion over such wide-field images is represented by the spatially variant PSF (SVPSF). The spatially invariant PSF (SIPSF) is used to define distortion over an isoplanatic region. However, distortion over anisoplanatic regions, i.e., over a wide, spatial field, can be represented using multiple SIPSFs. Block processing is used to combine each restored isoplanatic region. Restoration of wide-field images is performed using a block algorithm, where individual distortion operators are used to restore isoplanatic portions of a complete anisotropic image.

Using the continuous SVPSF, the MSE was calculated over each isoplanatic region and was shown to increase in moving from the centre of each region, reaching a maximum at each isoplanatic boundary. A restored, anisotropic image was compared with the original image using both the MSE and similarity metric. Block boundary errors were reduced using a linear filter applied over a portion of each boundary. This method employed an interpolated SIPSF, representing the distortion function between boundaries. Further deconvolution, using an interpolated SIPSF from adjacent regions, formed a *patch* that was applied over each boundary. This enhanced block processing method was applied to a high-resolution image of the Moon, were several SIPSFs were simulated and convolved with the original image to test the theory. The results, in terms of two quality metrics, are presented and briefly discussed. Lastly, a summary of the restoration methods discussed in this appendix is given.

A.1 Introduction

The restoration of astronomical images is an important affirmation of this research. However, knowledge *a posteriori* of the SVPSF is a prerequisite to minimise error between a diffraction limited image of an object, and the restored image of an object distorted by wavefront aberration. Alternative restorative methods that do not employ PSF estimates from reference beacons, such as blind deconvolution [7], use an iterative, maximum likelihood (ML) approach to determine the distortion function. While this is an effective method used widely for astronomical image restoration, the algorithm is based on noise statistics *a priori*, and is generally applied when the PSF is space invariant; a notable exception is a space-variant, neural network approach proposed by Cheema *et al.* [18].

A.2 Degrees of freedom

Degrees of freedom (DOF) is an interdisciplinary term that quantifies a minimum dependence on parameters to support the operational requirements of a system. Formally, DOF has been defined as, "the difference in dimensionalities of parameter spaces" [41]. In terms of imaging, Andrews and Hunt state that DOF is a concept used to describe an image with the minimal number of independent parameters [2]. This section defines the degrees of freedom used in anisoplanatic imaging by qualifying block size of the SIPSF prior to deconvolution.

Four 1D examples of DOF are shown in Figure A.1. Subfigures (a) and (b) employ the SIPSF, however insufficient sampling used in (b) introduces correlation between the first and third sample points, i.e., $R_{13} \neq 0$; such conditions typically result in poor imaging. Ideally, sample points should be placed such that minimal correlation exists between adjacent samples; this is shown in Subfigure (a).

For anisoplanatically blurred images, the SVPSF is used to represent variations in distortion over the image plane. This is shown in Subfigure A.1 (c) using fixed sampling, and (d) using variable spatial sampling. In the case of Subfigure (c), an increasing correlation between sample points 2, 3, and 4, i.e., $R_{12} < R_{23} < R_{34}$, results in distortion between adjacent samples. This problem is addressed in Subfigure (d), where variable sampling is applied. However, the implementation of variable sampling is non-trivial. In addition, each restored anisoplanatic image would require a different set of sample points. Knowledge *a priori* is needed to implement such a configuration.

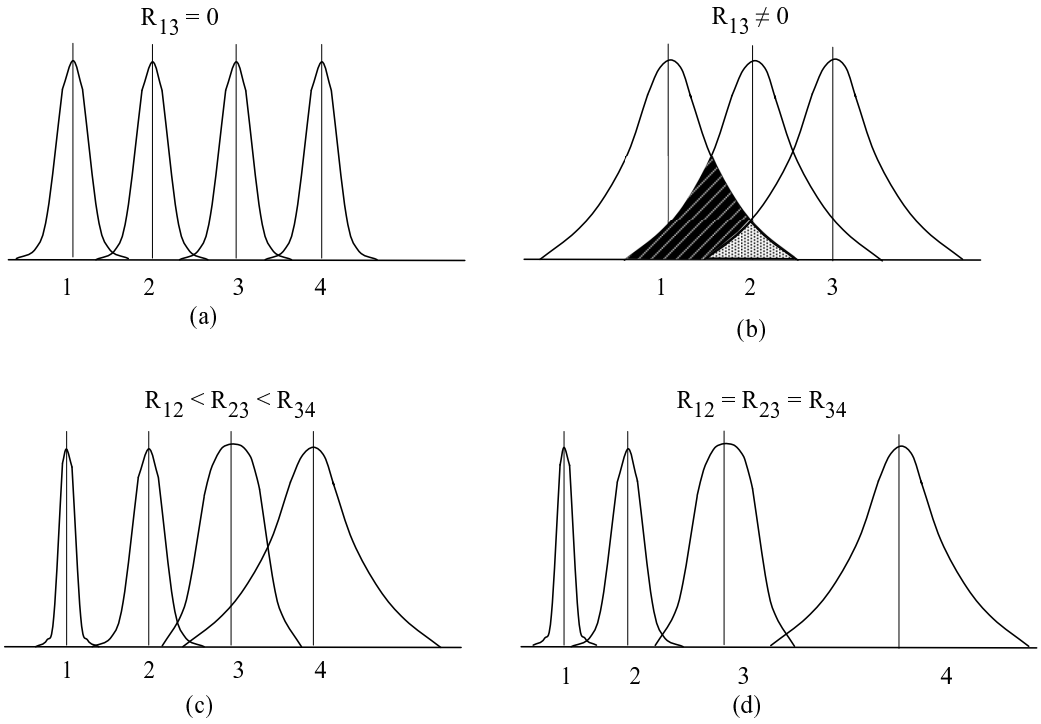

Figure A.1 Degrees of Freedom - spatial considerations: (a) SIPSF (good imaging); (b) SIPSF (poor imaging); (c) SVPSF (fixed sampling); (d) SVPSF (variable sampling). Adapted from Andrews and Hunt [2].

This issue is partially addressed by using fixed spatial sampling over designated regions; such regions are referred to as *blocks* [4], and their application is referred to as *block processing* [6]. Related issues concerning the size of each block, and distortion that can result between adjoining blocks, are addressed in the reminder of this appendix.

A.2.1 PSF spatial constraints

In terms of this study, DOF defines the number of independent distortion functions, $h(\cdot)$, required to restore an anisoplanatically distorted image, where $h(\cdot)$ can be either spatially variant or spatially invariant. Although related to the Nyquist criterion that defines a minimum spatial frequency to avoid aliasing, the DOF is used in this context to define the applicability of each SIPSF within an image, in terms of block size.

For anisoplanatic imaging, the SVPSF is employed. Fixed spatial sampling is used, and block size is based on minimal PSF variation over each isoplanatic region, i.e., it is based on the isoplanatic angle θ_0, as defined by Equation 4.16.

As a consequence of Fourier-based deconvolution processes, boundary artefact distortion is generated around each block due to segment effects [2]. A method proposed by Aubailly and Roggemann *et al.* [6] was used. This method used only the inner portion of the restored image, where artefact distortion contained in around the outer portion of restored image is omitted. An example of this technique, where the inner portions of two overlapping $\theta_0 \times \theta_0$ blocks are used for restoration, is shown in Figure A.2.

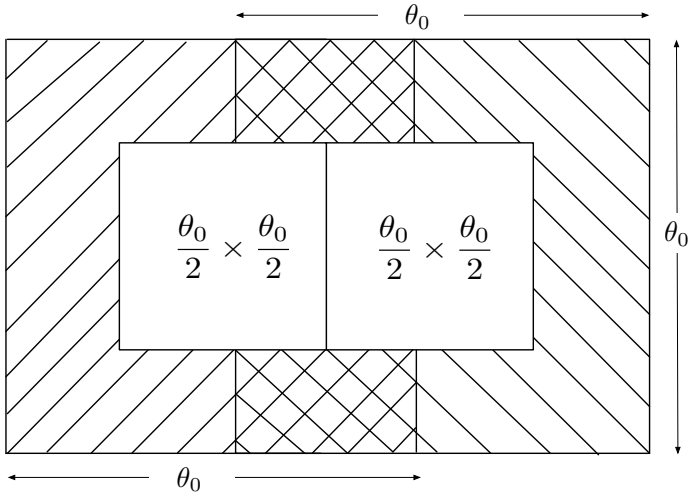

Figure A.2 Block restoration method to minimise boundary artefacts in anisoplanatic image restoration. Adapted from Aubailly and Roggemann [6].

The aforementioned method was used to effectively restore distorted images of extended objects, however early results suggested minor boundary distortion was still present. Close inspection of Figure A.2 shows the affected region as the line adjoining each pair of adjacent $\theta_0/2 \times \theta_0/2$ sized blocks. A 1D example of this configuration in Figure A.3.

A method proposed by Andrews and Hunt [2] was used to estimate the set of distortion functions (SVPSFs) comprising an anisoplanatic image. Specifically, the adjoining regions were examined using a correlation matrix, R_{ij}, for SIPSFs from adjoining regions. This correlation function is defined as [2]

$$R_{ij} = \int_{\infty}^{\infty} \int_{\infty}^{\infty} h_i(\xi,\eta)\, h_j(\xi,\eta)\, g\xi d\eta, \tag{A.1}$$

where $i, j = 1, \cdots, N$, for an $N \times N$ sized correlation matrix.

Figure A.3 shows two SIPSF sets, $\{h_1(\xi)\}$ and $\{h_2(\xi)\}$, where the DOF is defined to minimise correlation between each sample, and SIPSF block size is determined by θ_0. Deconvolution over each associated block is required but is constrained by θ_0. For example,

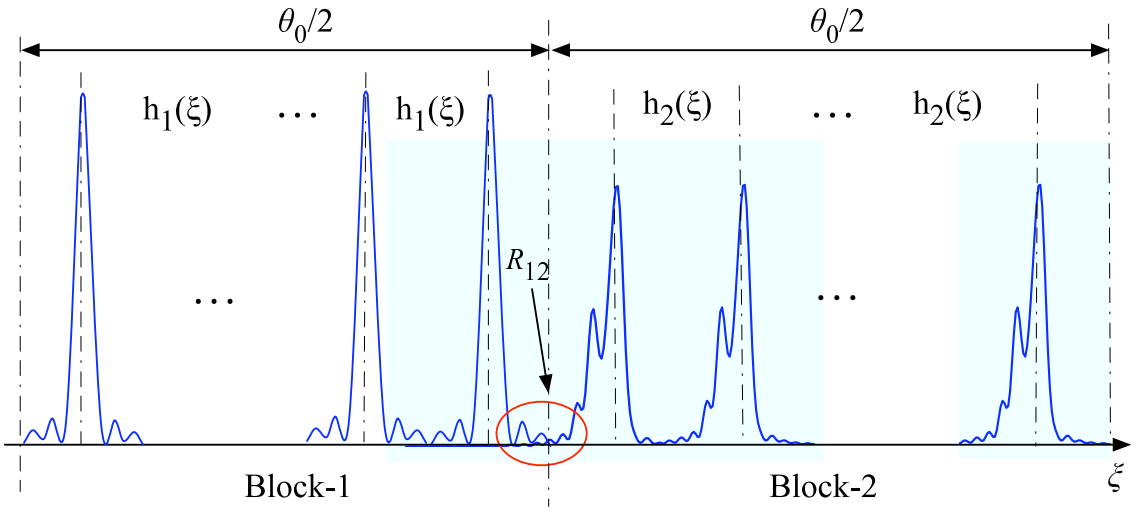

Figure A.3 SVPSF spatial considerations for anisoplanatic imaging.

$h_1(\xi)$ is used to deconvolve a distorted image portion defined over Block-1; a separate deconvolution is required over Block-2 using the SIPSF, $h_2(\xi)$.

Over each spatially sampled block, the assumed SIPSF remains constant. However, since the distortion function in object space is continuous, variations within SIPSF blocks will occur. An error, in terms of MSE, is defined as the mean squared difference between the SIPSF and the continuous distortion function. Such an error will gradually increase from the centre of each block and reach a maximum at each border; this is highlighted in Figure A.3 as R_{12}. Distortion resulting from this error can be represented by off-diagonal elements of the correlation matrix.

Thus, the SVPSF is defined in terms of blocks, where each block represents a SIPSF over a region defined by θ_0. Undesirable artefacts resulting from deconvolution are minimised by employing overlapping blocks, resulting in the restoration of $\theta_0/2 \times \theta_0/2$ sub-images. However, the effect of boundary distortion is non-trivial and required further investigation.

A.3 PSF error maps

In this section the extent of SVPSF evolution, as it explores an anisoplanaticlly distorted image, is determined. A PSF error map was constructed using the MSE metric; this is defined as the mean squared error between a reference PSF, positioned at the centre of a

isoplanatic block, and the continuous SVPSF, as the PSF evolves between blocks. To perform this analysis, multiple PSFs were required and a mathematical construct was needed to generate MATLAB® code. A construct for N blocks can be defined as

$$\varepsilon^2(x,y) = \bigcup_{n=1}^{N} \Big\{ \varepsilon_1^2\left[h_{\mathrm{Ref}_1}(x,y), h(x,y)\right], \varepsilon_2^2\left[h_{\mathrm{Ref}_2}(x,y), h(x,y)\right], \cdots \\ \varepsilon_n^2\left[h_{\mathrm{Ref}_k}(x,y), h(x,y)\right] \Big\}, \quad \forall k \in \{1,2,\cdots,N\}, \tag{A.2}$$

where $\varepsilon_k^2[\cdot]$ is the block MSE given by Equation 4.17, h_{Ref_k} can be either the k^{th} point source beacon, S_k, or target object, T_k, $h(x,y)$ is the SVPSF over (x,y) and where $x,y \in$ block k.

A distinction can be made between a wavefront map and a PSF error map. The former is generated by propagating one or more source beacons, $S_1, S_2, \cdots, S_N$ through a phase screen, employing the simulation environment described in Subsection 6.2.4. However the latter is generated by applying Equation A.2 and calculating the residual, PSF error, as the distortion function explores image space. An example of an error plot for a single source object is shown in Figure A.4.

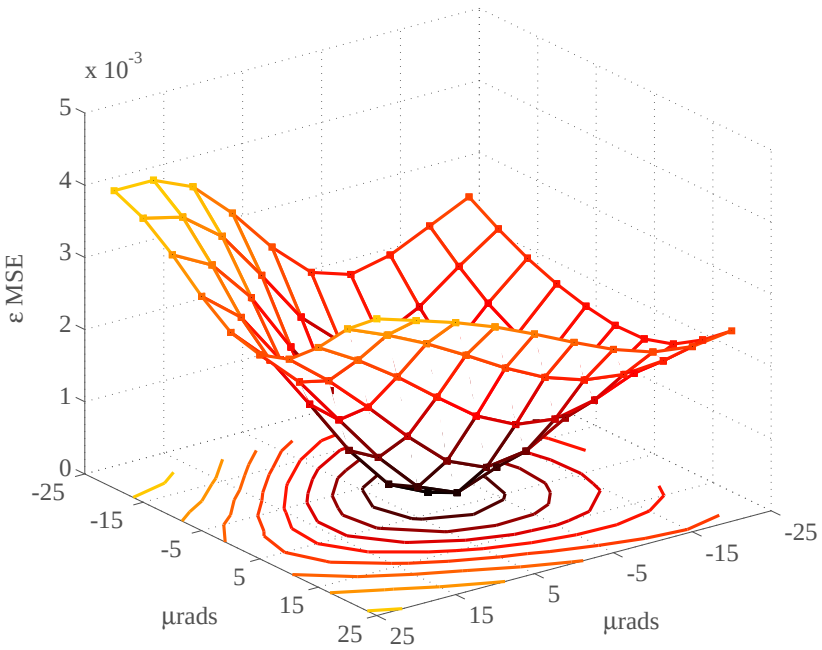

Figure A.4 Error plot of the spatially variant PSF for a single source object.

Ideally, for restoration of extended objects, a SIPSF should be defined for each spatial sample defined on the image plane. However, computationally, this is not practical. The ESN method can be used to predict the SVPSF, however predicting multiple SIPSFs at a system sampling rate of 60 Hz is not possible; estimating a single, point source target object over relatively wide field angles is achievable, given intensive ESN training prior to testing.

A.4 Open loop imaging model

An open loop imaging model is defined in this section, and is based on the system model shown in Figure 6.1. However, this model includes measurement noise and verification ports, outlined in Figure 6.12. The model can be described as follows. Modal wavefront data from source beacons, $\mathbf{z}_K$ are used with angular separations between source beacons and target objects, $\boldsymbol{\theta}_{\text{Sep}}$, to predict modal wavefront aberrations of targets, $\hat{\mathbf{z}}_{\mathbf{T}}$. From these estimations, the residual wavefront phase error, $\varepsilon_R^2(D/r_0)$, is calculated using actual, modal wavefront data, $\mathbf{z}_{\mathbf{T}}$.

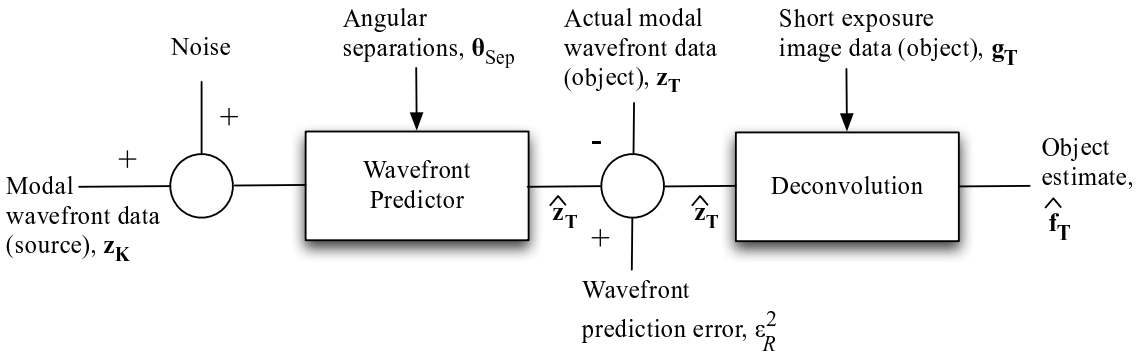

Figure A.5 Open loop imaging model.

The second stage of the model can be described as follows. Estimated wavefront data, $\hat{\mathbf{z}}_{\mathbf{T}}$, is used to reconstruct the SVPSF. The estimated SVPSF, in conjunction with the distorted image, $\mathbf{g}_{\mathbf{T}}$ are then deconvolved using a suitable algorithm, to recover an estimate of the object, $\hat{\mathbf{f}}_{\mathbf{T}}$.

The open loop model is based on deconvolution from wavefront sensing (DWFS), originally proposed by Primot *et al.* [106]. The DWFS method was extended by Roggermann *et*

al. [116] to widen the FOV of optical telescopes. However, the ESN method, defined as the *wavefront predictor* in Figure A.5, proposes a different approach to estimate the SVPSF. Given knowledge *a posteriori* of the SVPSF, the final process required to restore anisoplanaticlly blurred images is detailed in the next section.

A.5 Image restoration

In this section, two widely used image restoration algorithms are described. A complete anisotropic image is defined that can be generated using the block processing method. A modified version of the block processing method is presented that further reduces boundary distortion. Lastly, the block processing method is described to provide insight into the application of the SIPSF.

A.5.1 Restoration algorithms

The forward image model is used to define an image, $g(x,y)$, as the convolution between an object, $f(x,y)$, and the PSF, $h(\cdot)$. This can be represented in matrix form as

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\eta}, \tag{A.3}$$

where $\mathbf{H}$ is the matrix formed from the PSF, when $\boldsymbol{g}$ and $\boldsymbol{f}$ are vectors formed from $g(\cdot)$ and $f(\cdot)$, taking pixels in raster order, and $\boldsymbol{\eta}$ is noise.

Image restoration is an example of an ill-posed problem [50]. A simple inversion of $\mathbf{H}$ in Equation A.3 results in the amplification of noise. To reduce the effects of this undesirable result, regularisation is used. Two, well-know methods used for astronomical image restoration include Tikhonov regularisation and the expectation maximisation (EM) algorithm.

Tikhonov regularisation defines an energy constraint, or degree of smoothness for the solution, and has been applied to shift-varying degradation problems [9]. Using such a constraint, an estimate of the original image can be obtained by

$$\hat{\mathbf{f}} = [\mathbf{H}^T\mathbf{H} + \alpha\mathbf{C}^T\mathbf{C}]^{-1}(\mathbf{H}^T\mathbf{g}), \tag{A.4}$$

where $\mathbf{C}$ is the 2D Laplacian operator and α represents the Lagrange multiplier, also known as the regularisation parameter [8].

An alternate restoration method is the expectation maximisation (EM) algorithm. This is also known as the Lucy-Richardson (LR) algorithm [23]. The EM algorithm uses the normalised PSF as a probability density function.

Given a Poisson noise distribution, the EM algorithm provides a maximum-likelihood solution. However, the EM algorithm can still be applied in the absence of photon noise. A recursive form of the EM algorithm is given by [93]

$$\hat{\mathbf{f}}_i^k = \hat{\mathbf{f}}_i^{k-1}\left(\frac{1}{\sum_j \mathrm{H}_{ij}}\right)\sum_j\left(\mathrm{H}_{ij}\cdot\frac{\mathbf{g}_j}{\sum_l \mathrm{H}_{jl}\hat{\mathbf{f}}_l^{k-1}}\right), \tag{A.5}$$

where $i \in \{1,2,\cdots,N^2\}$, and H_{ij} is the i^{th} element of the j^{th} column of matrix H.

An element, (i,j) of H, has the probability that a photon from the i^{th} pixel location in the object plane will fall into the j^{th} pixel location in the image plane.

The LR-EM algorithm was used in this research to restore sub-images that represented isoplanatic regions.

A.5.2 Isoplanatic blocks

An introductory discussion on image applications, where the PSF can be described as either spatially variant or invariant, is given in Section 3.7. In this section, this qualification is extended, and the constraints used to define spatial invariance are addressed.

Spatial invariance of the PSF is a defining requirement for isoplanatic regions, i.e., when the linear operation of deconvolution is applied, the SIPSF is held constant over a region assumed to be isoplanatic. In practice however, minor perturbations must be tolerated. Alternatively, every pixel in the resulting image would require support of an individual SIPSF.

For computational convenience, deconvolution algorithms were applied to rectangular blocks. The SIPSF can be defined by a Toeplitz matrix [2]; the application of Toeplitz matrices is discussed in Subsection A.5.4.

Consider an isoplanatic block, Λ. The recorded image in the block, $g_\Lambda(x,y)$, can be described as

$$g_\Lambda(x,y) = \sum_{k,l\in\Lambda} f_\Lambda(k,l)h_\Lambda(x-k;y-l) + \eta(x,y), \tag{A.6}$$

where $f_\Lambda(\cdot)$ is a distorted image bound by Λ, $h_\Lambda(\cdot)$ is the SIPSF applicable to Λ, and $\eta(x,y)$ is observation noise, comprising both read noise with zero mean and variance σ^2, and photon noise.

The restoration of anisoplanatically distorted images requires the estimation of the SVPSF, in terms of a set of SIPSFs. A process is developed for the anisoplanatic model, and is described in the following section.

A.5.3 The anisotropic image

Application of a deconvolution algorithm, such as Equation A.4 or A.5, on a given patch of $g(x,y)$, using the SIPSF, $h_\Lambda(x,y)$, results in an estimated image portion of the original object, $\hat{f}_\Lambda(x,y)$. These portions are represented as square block matrices. To simplify notation, Λ is used to represent a portion of the restored image, $\hat{f}(x,y)$, and matrix notation, e.g., $\Lambda_{i,j}$, is used to distinguish the position of each restored block within the reconstructed image.

An example of a restored anisoplanatic image, $\hat{f}(x,y)$, arranged as a 2×2 isoplanatic block matrix, is shown in Figure A.6.

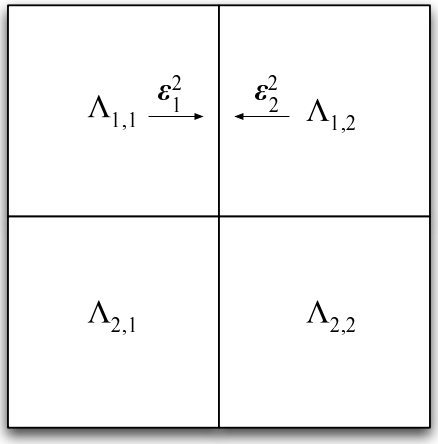

Figure A.6 Isoplanatic regions and SIPSF boundaries.

Consider a single deconvolved block, labelled $\Lambda_{1,1}$ in Figure A.6, represented by a single SIPSF at the centre of the region. Since the SVPSF is continuous, as the distortion function migrates from block $\Lambda_{1,1}$ to $\Lambda_{1,2}$, the MSE error, shown as ε_1^2 is expected to increase

towards the boundary. A similar situation exists for block $\Lambda_{1,2}$, where the error, ε_2^2, is also expected to increase spatially in the direction of block $\Lambda_{1,1}$. The result is a small, but measurable discontinuity of the SVPSF at the boundary of these regions. This is also represented by the correlation that exists between each adjacent PSF, R_{12}, as shown in Figure A.3.

Now consider a modification to Figure A.6, where each adjacent SIPSF is interpolated, resulting in four, additional PSFs. The interpolated PSFs comprise an average of each aberration used to construct each adjacent PSF. These interpolated PSFs were used to deconvolve a smaller region (a sub-block) over each boundary. The orientation of each sub-block is shown in Figure A.7 as, S_1, S_2, S_3, and S_4.

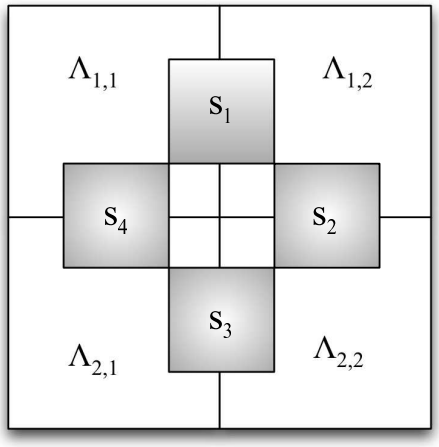

Figure A.7 Orientation of sub-blocks on SIPSF boundaries.

Thus, a complete anisotropic image was reconstructed using a *tiled* or *mosaic* approach [28], and where $\theta_0/2 \times \theta_0/2$ sized blocks were interpolated to reduce SVPSF discontinuities that exist across the boundaries of each block [143]. Incorporation of this technique with a block processing method was performed; this is discussed in the following subsection.

A.5.4 Block processing

To facilitate the reassembly of isoplanatic blocks to form a complete anisotropic image, an algorithm proposed by Aubailly was used [3]. The algorithm firstly extracts a $\theta_0 \times \theta_0$ sized block from an anisoplanatic image and this is deconvolved using the predicted SIPSF;

details of this process are described in Section A.2. The restored block is then reassembled as part of the anisotropic image. The algorithm iteratively processes N blocks until the complete anisotropic image has been restored.

The method outlined in the preceding subsection was integrated with the block processing algorithm proposed by Aubailly [3]. A new algorithm incorporating both methods is described in the flow-chart shown in Figure A.8.

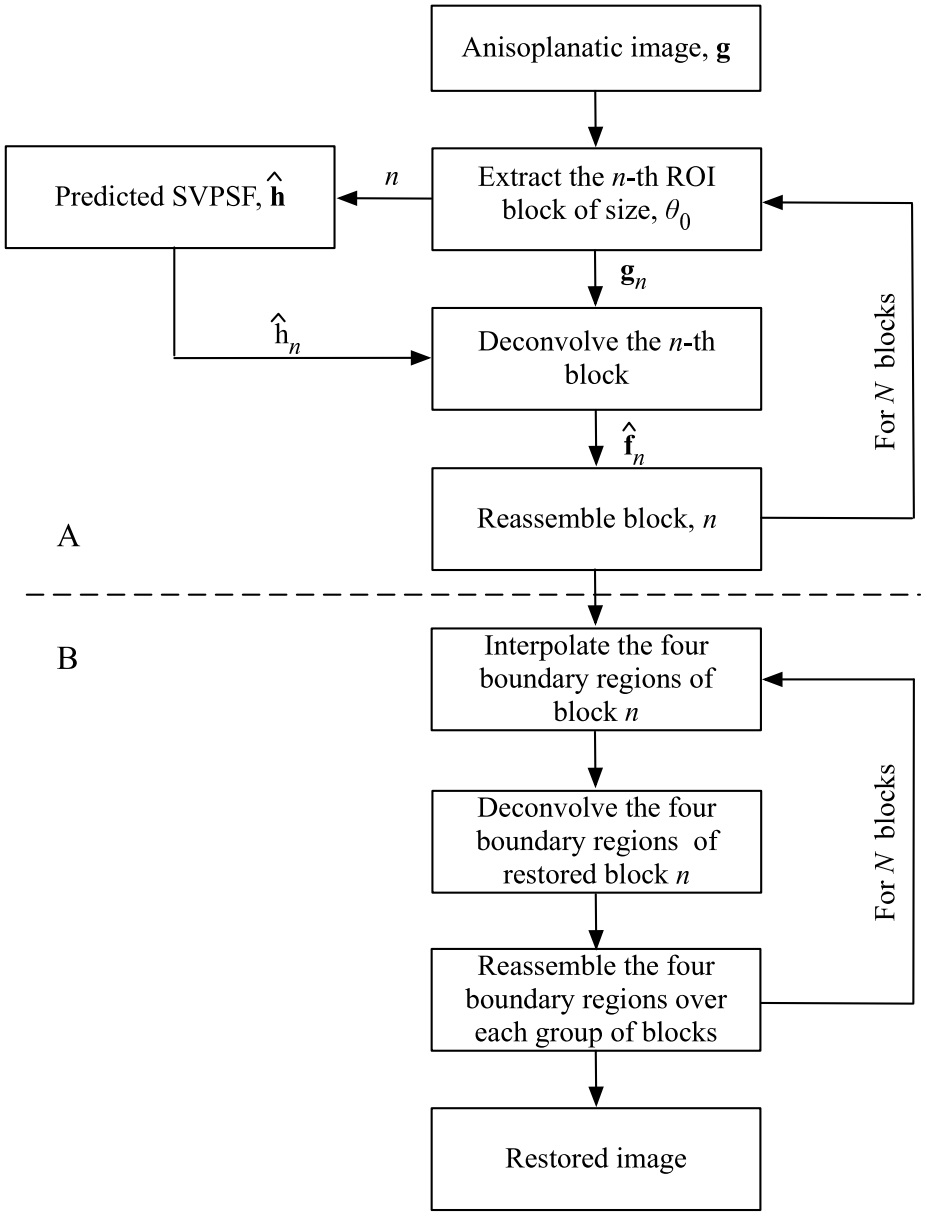

Figure A.8 Combined block processing and PSF interpolation method.

The algorithm described in Figure A.8 is separated into two parts. The upper section, Part-A, incorporates an existing block processing algorithm by Aubailly [3]. However, the method used by Aubailly to predict the SVPSF differs from the ESN method described in this thesis. The former is summarised in Section 4.6, whereas the latter is the principal topic discussed throughout this thesis. The lower section, Part-B, shows the proposed PSF interpolation method described in the preceding subsection.

Real-time image processing requires efficient methods for image restoration. Although not specifically used in this research, such methods are described in this section as an extension to this discussion on block processing.

Characteristics regarding the cyclic nature of convolution, and the efficiency of the discrete Fourier transform, can be employed by preprocessing data for use in efficient algorithms. For example, each $\theta_0 \times \theta_0$ sized block from the distorted image is spatially invariant. As such, $\mathbf{H}$, shown in Equation A.3, is in Toepliz form [46]. Matrix forms of the distorted image, $\mathbf{g}$, original image, $\mathbf{f}$, and noise, $\boldsymbol{\eta}$, defined in Equation A.3, are represented by $N^2 \times 1$ vectors, $\mathbf{H}$ is an $N^2 \times N^2$ matrix and can be approximated by a *block circulant* matrix [2]

$$\mathbf{H}_{\mathrm{BC}} = \begin{bmatrix} \mathbf{H}_0^c & \mathbf{H}_{N-1}^c & \mathbf{H}_{N-2}^c & \cdots & \mathbf{H}_1^c \\ \mathbf{H}_1^c & \mathbf{H}_0^c & \mathbf{H}_{N-1}^c & \cdots & \mathbf{H}_2^c \\ \mathbf{H}_2^c & \mathbf{H}_1^c & \mathbf{H}_0^c & \cdots & \mathbf{H}_3^c \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{N-1}^c & \mathbf{H}_{N-2}^c & \mathbf{H}_{N-3}^c & \cdots & \mathbf{H}_0^c \end{bmatrix}, \tag{A.7}$$

with each partition in $\mathbf{H}_{\mathrm{BC}}$ a circulant matrix, $\mathbf{H}_j^c$, [2]

$$\mathbf{H}_j^c = \begin{bmatrix} h(j,0) & h(j,N-1) & h(j,N-2) & \cdots & h(j,1) \\ h(j,1) & h(j,0) & h(j,N-1) & \cdots & h(j,2) \\ h(j,2) & h(j,1) & h(j,0) & \cdots & h(j,3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h(j,N-1) & h(j,N-2) & h(j,N-3) & \cdots & h(j,0) \end{bmatrix}. \tag{A.8}$$

The significance of block circulant matrices for space invariant restoration is that they are easily diagonalised; their eigenvalues are the 2D DFT values of the defining 2D sequences,

and their eigenvectors are defined in terms of Fourier kernels [68, 69].

A.6 Restoration example

A restoration example using a high-resolution image of an extended object is presented in this section. The purpose of this section is to demonstrate the application of several restorative algorithms described in this appendix, and assesses the SVPSF interpolation method proposed in Section A.5. A high-resolution image of the Moon was used for this example [100].

A.6.1 Procedure

Simulations were performed using the framework discussed in this thesis to generate several SIPSFs. Each SIPSF was used to distort a quadrant of a high-resolution image of the surface of the Moon. The original, undistorted image is shown in Subfigure A.10 (a).

A set of Zernike coefficients were used to construct a series of SIPSFs using Equation 3.41; each SIPSF represented an isoplanatic region. A a set of four SIPSFs were used to represent the spatially variant PSF. The distortion functions and their configuration used to generate an anisoplanatic image is shown in Figure A.9. Three of the four SIPSFs produced motion blur. The fourth SIPSF was configured to induce a Gaussian degradation over the low-left corner of the image.

Using each SIPSF, a linear convolution was performed on a respective portion of the image. For example, each blur function aberrated the image in the direction of the SPISF. This resulted in the *smearing* of craters in the direction of the distortion function. However, the Gaussian distortion was similar to defocus, where the edges of craters were lost completely in the application of this distortion function.

Each distorted image was reassembled to form a complete, anisotropic image. In addition, read noise was uniformly added over the entire image, such that SNR $\approx 40\,\text{dB}$. The distorted image is shown in Subfigure A.10 (b).

A.6.2 Restoration

Block processing was used to restore each isoplanatic sub-image. The procedure outlined in Part-A of Figure A.8 was adopted. This process resulted in four, restored sub-images.

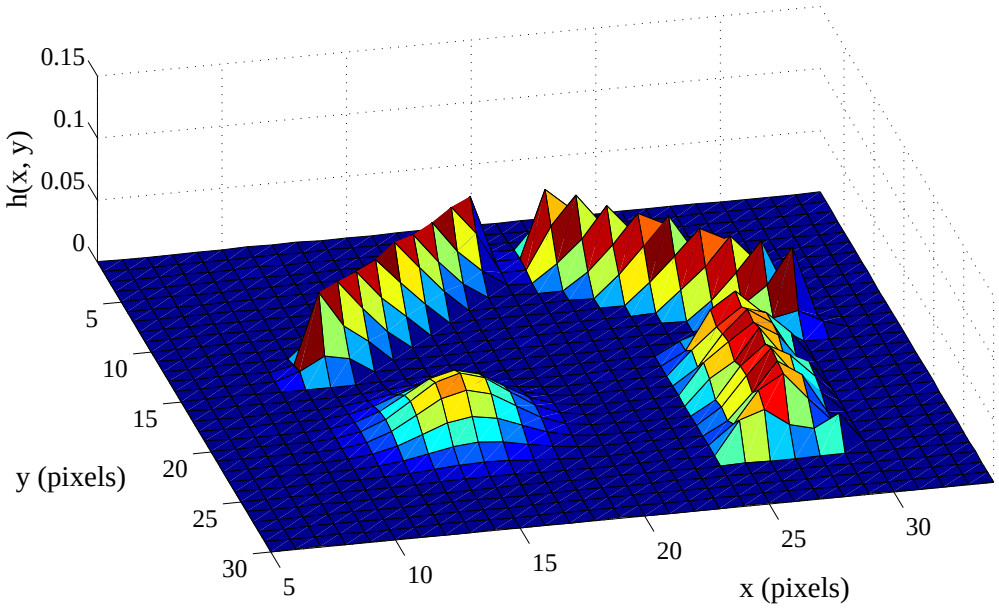

Figure A.9 The spatially-variant PSF.

Each sub-image was repositioned as a mosaic; the resulting image is shown in Subfigure A.10 (c).

To evaluate the effectiveness of minimising boundary distortion, a set of Zernike coefficients, $\{a_2, a_3, \cdots, a_{10}\}$, composed from each adjacent SIPSF used to distort portions of the original image, were linearly interpolated to construct a new SIPSF. The LR-EM algorithm was then used to reconstruct the boundary portion between each sub-image. Each boundary region corresponded that those shown in Figure A.7, labelled, S_1, S_2, S_3, and S_4. Each sub-block effectively replaced each boundary region to minimise discontinuities between isoplanatic regions. The resulting image is shown in Figure A.10 (d).

A visual inspection and comparison of Subfigures (c) and (d) in Figure A.10 shows marginal improvement in the four surrounding border regions. Since a visual comparison of these images is difficult to assess, a quality analysis was performed. The results of this analysis are presented in the following subsection.

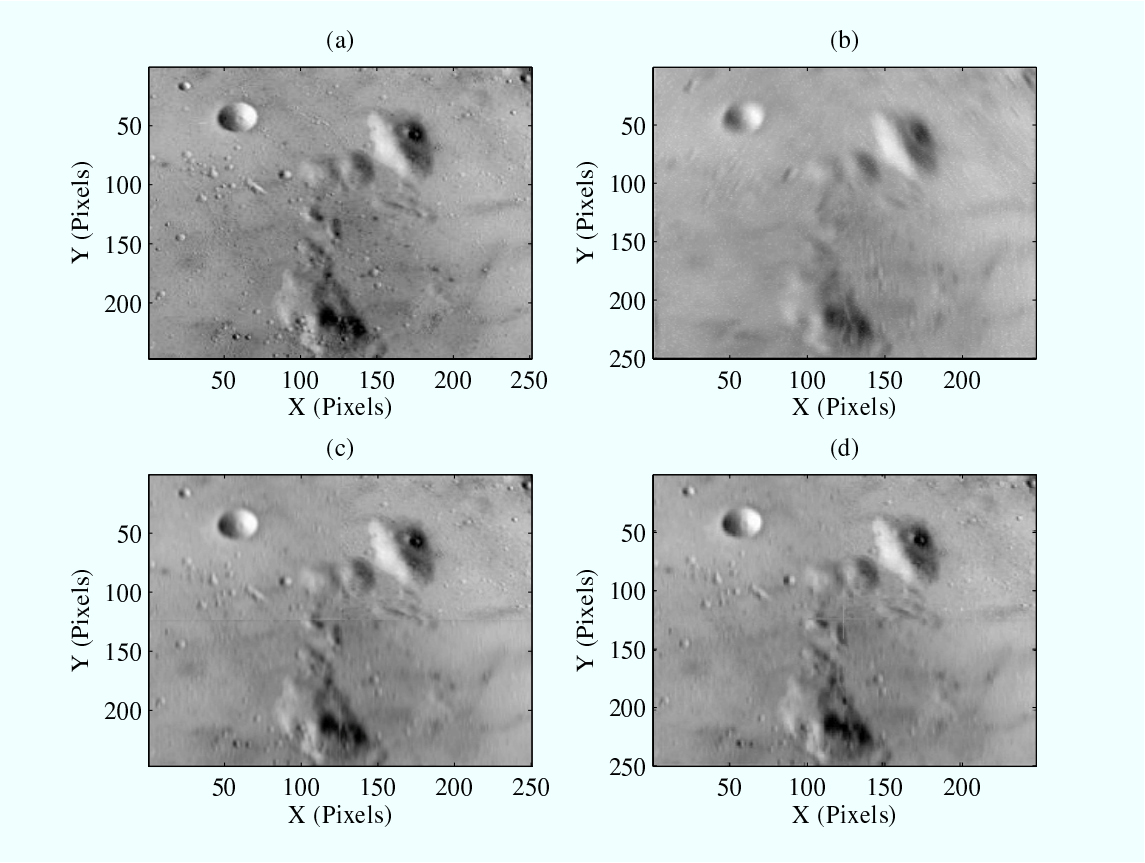

Figure A.10 (a) Original extended object, (b) anisoplanatically distorted image, (c) deconvolved image using four SIPSFs, (d) deconvolved image with linear filtering.

A.6.3 Results

A metric is required to assess the quality of restorations. The mean-squared error (MSE) metric is often used in image restoration to compare the accuracy of restorations [4], however a quality measure was also used in this analysis. This quality measure is referred to in the literature as the *similarity* metric [139]. The similarity metric (SSIM) is a method that can be applied to assess the perceptual qualities of an image, and is defined by Equation 2.22.

The results of a comparison between a restored and reassembled anisotropic image, with and without SVPSF interpolation, to the original undistorted image, is listed in Table A.1. Both the MSE and SSIM metrics were used in this comparison, since both are quality assurance measures.

Only a portion of each of each block boundary was restored using the process outlined in Part-B of Figure A.8. For example, the central sub-block, surrounded by sub-blocks S_1, S_2,

Table A.1 Image Reconstruction Comparisons

Image Comparison of Figure A.10	Metric	
	MSE	*SSIM*
Tiled SIPSF, (a) & (c)	14.54	0.7125
Interpolated SIPSF, (a) & (d)	13.86	0.8546

S_3, and S_4, was not processed using the interpolated PSF method. However, improvement along a portion of the four borders shown in Subfigure (d), is clearly evident. This method effectively removed border distortion by deconvolving these regions with an interpolated PSF.

Lastly, the MSE and SSIM metrics were used to compare the reconstructed image in Subfigure A.10 (c), with the highlighted portion of the original image shown in Figure A.10 (a). The original image was then compared to a restored anisotropic image shown in Figure A.10 (d). The results listed in Table A.1 show a small improvement, i.e., a reduction in terms of MSE. This process was repeated using the SSIM to assess the perceptual quality of the restoration. The SSIM assessment shows a considerable improvement of $\approx 14\%$, where a perfect match is equated to unity.

A.7 Summary

In summary, this appendix has described several processes used in the restoration of astronomical images. More specifically, various methods utilising the SVPSF as a set SIPSFs, were used to restore a complete anisotropic image. Image reconstruction is based on the reassembly of each restored block. A principal focus in this appendix is on boundary distortion, evident at the borders of each block resulting from small errors that exist over isoplanatic blocks.

To supplement the ESN method for estimating the SVPSF by predicting wavefront aberrations over a wide FPV, the PSF interpolation can be used in restoration to reduce block boundary distortion. A high resolution image of the Moon was used to test this restoration method. The MSE and SSIM metrics were used to compare the original image with reassembled and restored, anisotropic images. A set of these images including the SVPSF

interpolated image. The results using both metrics showed improvement when the SVPSF interpolation method was employed.

Appendix B

Laboratory Equipment and Experimentation

In this chapter an experimental platform is described for acquisition of empirical data; this platform is offered as a possible alternative to data acquired in the field. The experimental platform was established in an Optics Laboratory, supported by the Department of Electrical and Computer Engineering at the University of Canterbury, Christchurch, New Zealand. The set-up and configuration of equipment used for this platform is detailed in this appendix.

The organisation of this appendix is as follows. A description of equipment used in this study, including the adaptation of an optical breadboard, designed and built from collaboration between the departments of Physics and Astronomy and Electrical and Computer Engineering at the University of Canterbury [63], is detailed in Section B.1. This is followed by a detailed description of an experimental platform, designed for use in an Optics laboratory, in Section B.2.

B.1 Equipment

The equipment described in this section was used to acquire data for validation of the methods detailed in this dissertation. A principal component shared between both observational and laboratory data acquisition was an optical breadboard. Originally designed for SCIDAR observation runs [63], the configuration of the rig was modified to support two, high-speed CCD cameras for wavefront sensing. A third, in-focus CCD camera is used for DWFS. De-

tails on supplementary optical components, used in conjunction with the optical breadboard in the laboratory environment, are provided in Subsection B.2.1.

The optical breadboard shown in Figure B.1 supports three, high-speed CCD cameras from Point Grey Research[1]. Two cameras are placed at the intra- and extra-focal planes, as required by the extended geometric WFS. In Figure B.1 these WFS cameras are labelled `C2` and `C3`, respectively. A third, camera, labelled as `C3`, is placed at the focal plane and is used to acquire distorted images of astronomical objects for deconvolution from wavefront sensing (DWFS).

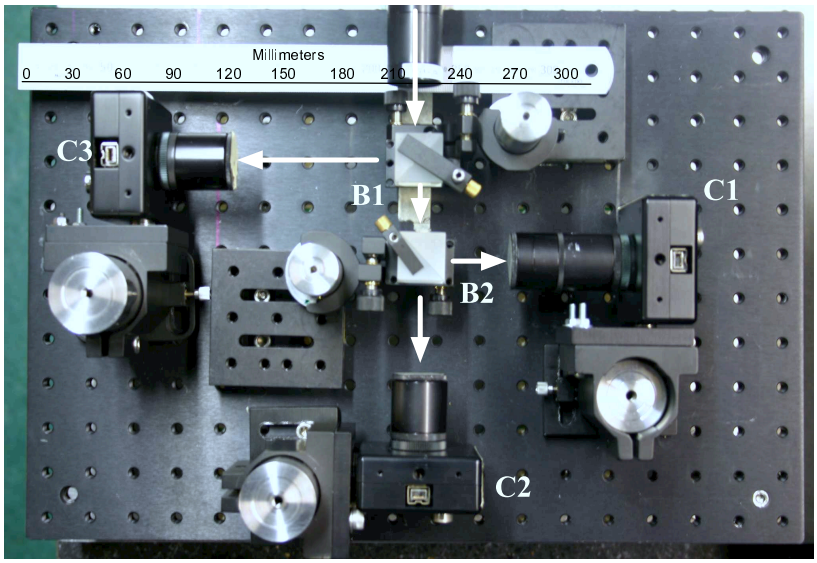

Figure B.1 Modified SCIDAR optical breadboard.

Two, 25.4 mm cube, 50/50 beamsplitters, constructed from BK7 glass with coatings for visual wavelengths 400 - 700 nm, are used on the optical breadboard. An arrow shown at the upper-central portion of Figure B.1 indicates the optical path from the exit pupil of the telescope. The optical path passes through the first beamspiltter, labelled `B1`. Light is equally distributed between camera `C3` and a second breamsplitter, labelled `B2` in Figure B.1. The beamsplitter labelled `B2` equally distributes incoming light from `B1` to wavefront sensing cameras labelled `C1` and `C2`. Unfortunately, beamsplitters exacerbate the problem of resolving images that are, in many cases, already photon starved[2]; their use reduces the number of stars which are candidates for source beacons. However, in contrast to a single

[1] model number DX-BW-CSBX

[2] An alternative is the use of photon-counting, avalanche photo diodes.

camera imaging system, such as employed by Ragazzoni *et al.* [110], scintillation effects are reduced using the Roddier (two camera) method [114]. Since wavefront amplitude fluctuations can have an adverse effect on phase measurements, the latter method was used for wavefront phase retrieval.

Each camera employs a Kodak KAI-0340 progressive scan CCD sensor, supporting a maximum framerate of 200 FPS. A Firewire, S800 IEEE 1394b interface is used to transfer data from each camera to a hard disk drive (HDD) via a host PC. However, due to data throughput limitations, both in terms of hardware and the supported operating system[3], it was not possible for two or more cameras to share the same PCI-e interface for frame rates in excess of 60 Hz. To ensure synchronisation of data, each camera supports an external trigger to initiate frame capture. To facilitate simultaneous frame capture, a waveform is generated and applied, in parallel, to the external triggers of each camera. Image capture is initiated on a falling (waveform) edge, and data from each camera are streamed to individual HDDs for post-processing. A purpose-built microcontroller module [140] is used to generate waveforms for trigger synchronisation, under the programmed control of a GUI. Both the camera and module have a latency of ≤ 5 ns. Data blocks from each camera are encoded to facilitate dual-frame sequencing for wavefront phase conversion. A frequency analysis, summarising the noise characteristics of the KAI-0340 CCD used in this study, is presented by Mohr [91].

A significant advantage of the SCIDAR optical breadboard is in the portability of the design, i.e., observational data can also be acquired from either the 1-m McLellan or the 60-cm Boller & Chivens telescopes at MJUO. In addition, laboratory equipment is compatible with the optical breadboard and is used to acquire data in either environment.

Lastly, the 1-m McLellan telescope can be configured to a focal ratio of either f/13 or f/7.7. Using the dimensions of the KAI-0340 image sensor, i.e., 7×5 mm, the FOV at prime focus is 480 μrador 910 μrad, respectively. The wavefront sensing cameras supported a sampling rate over 100 Hz. However, depending on seeing, and the objective of the experiment, most data acquisitions were conducted at 60 Hz.

[3] Only the Windows operating system was supported during the research period.

B.2 Laboratory

The department of Electrical and Computer Engineering at the University of Canterbury provided an Optical laboratory where experimentation on anisoplanatic imaging was conducted in a controlled environment. Unlike field observations, where photon starvation from faint astronomical objects was an ongoing concern, laser sources provided an abundance of photons resulting in minimal photon and CCD read noise.

Generating wavefront perturbations to accurately emulate the effects of atmospheric turbulence required specialised equipment [70, 71]. The incorporation of this equipment to emulate the effects of air turbulence on multiple source beacons is described in this section.

B.2.1 Test bench

Semiconductor lasers, of wavelength $\lambda = 630\,\text{nm}$, are used in the configuration shown in Figure B.2, providing four, planar wavefront sources. A turbulence generator was built to emulate phase perturbations, resulting from the propagation of one or more planar wavefronts through turbulence [129]. The Fried parameter, r_0, was approximated and remained relatively consistent using a closed-loop, microcomputer-based, control system. A modification to the turbulence generator was made, essentially channelling turbulent air to flow in a continuous, constant direction, i.e., emulating the displacement of a single *sheet* of turbulence across the aperture. This was required to establish a "frozen" turbulent flow, as proposed by the Taylor hypothesis.

The use of an aperture stop and optics provided emulation of a simple, single objective telescope. The telescope objective was adjusted to ensure that Fraunhofer and Fresnel diffraction patterns were measurable; the former was used for imaging focused point-sources, where deconvolution was applied to partially restore the image; the latter comprised a pair of defocused images for wavefront sensing. An optical breadboard, originally designed for SCIDAR [63] where a two camera system supported the acquisition of generalised and pupil plane images, was modified for use in this research.

B.2.2 Laboratory procedure

By employing the optical testbench in Figure B.2, three or more source beacons could be configured around a single, centrally located target. Each beacon could be moved independently over the exist pupil to form various field angle configurations.

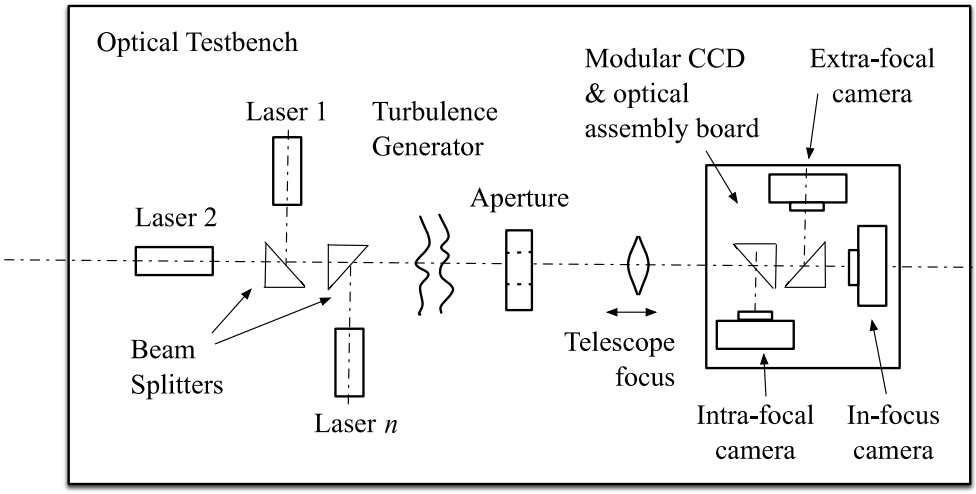

Figure B.2 Laboratory optical testbench configuration.

Four lasers were used to generate planar wavefronts based on the experimental platform outlined in Subsection B.2.1. A turbulence generator was used to perturb planar wavefronts [129], and a series of images were captured at 100 Hz using two CCD cameras. A set of frames representing intra- and extra-focal images is shown in Figure B.3. Each image is composed of four irradiance patterns and was defined by a region of interest (ROI). Batch processing using a WFS was required to extract individual wavefront maps from each perturbed source.

Wavefront maps from three source objects, S_1, S_2, and S_3, were used to estimate a single target, T. The extended geometric WFS was used to estimate four individual wavefront maps from eight ROIs, shown in Figure B.3 as T and T', and S_i and S_i', where $i \in \{1,2,3\}$. Each wavefront map comprised a Zernike modal set, and represented the distortion function from a corresponding source.

The optical testbench was used to acquire Zernike data ensembles for training the ESN. Time series data from each source, and angular separations from each source to a target object, were formatted and input to the ESN. Corresponding time series data from a single target facilitated ESN training. Several training sets, each using different angular separations and source configurations, were used to ensure network generalisation. Additionally, the turbulence generator could be programmed for various turbulence strengths. Calibrating each of these levels to a value of r_0 has been performed [71]. However, the classification method used for this testbench was based on a set of correlations performed on field data.

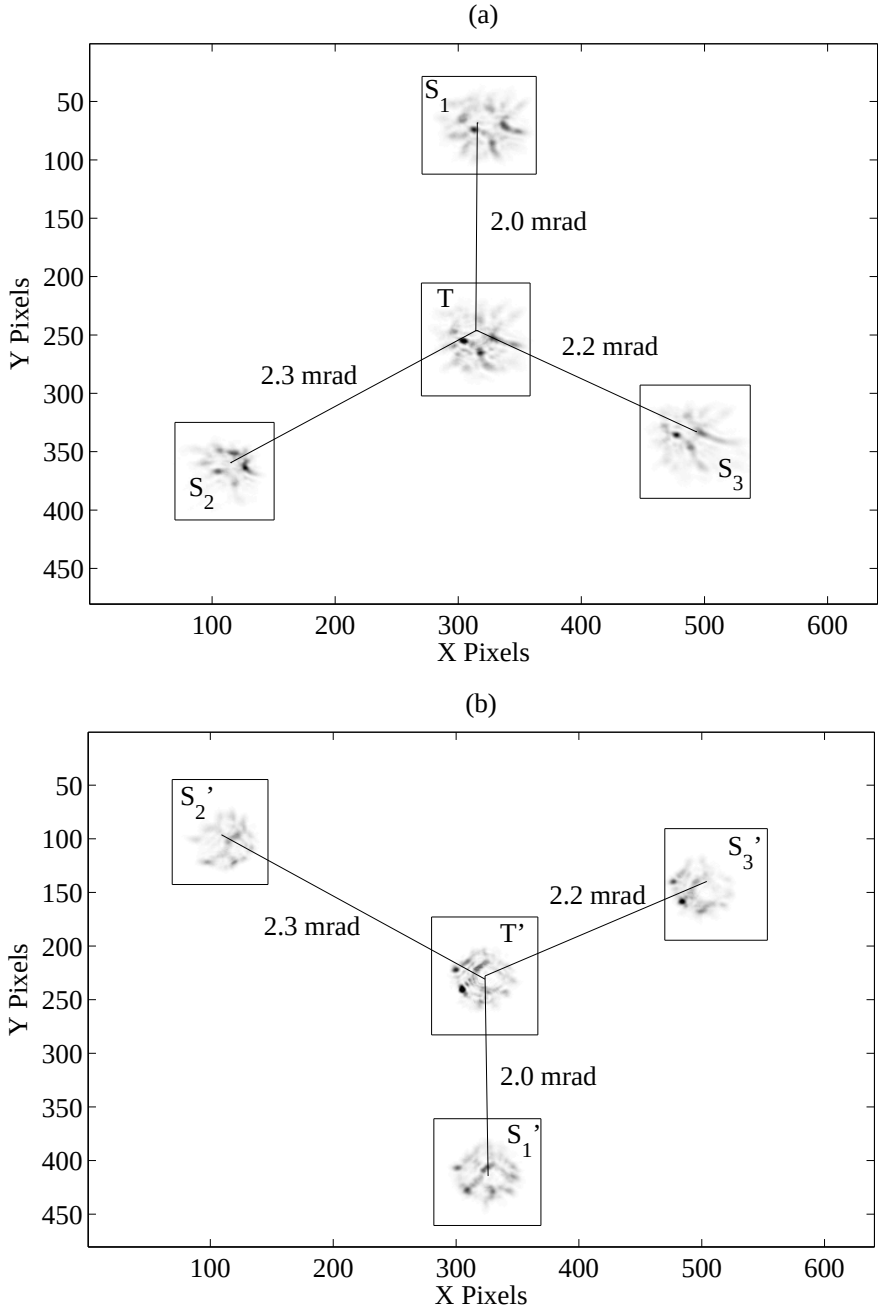

Figure B.3 Two synchronised frame captures representing intensity distributions of planar wavefronts generated from four artificial laser sources, and perturbed by a purpose-built turbulence generator: (a) inverted intra-focal image; (b) inverted extra-focal image.

MJUO turbulence profiles [91] provided the range of r_0 values used in this comparative analysis.

Lastly, the laboratory equipment and testbench configuration described in this appendix is at an intermediate stage. Additional equipment, such as a spatial light modulator (SLM) [111], is planned to replace the hot-air turbulence generator. Wavefront phase maps can then be generated using the methods discussed in Chapter 6; an embedded control system can be employed to move and update the phasescreen over the aperture. Individual phase measurements resulting from laser projections through the SLM are then recorded using high-speed CCD cameras discussed in the previous section.

Such extensions to current equipment, in addition to the configuration outlined in this appendix, are expected to simplify the acquisition of observational data for further research in anisoplanatic imaging.

Appendix C

Observational Data

The following pages provide details of empirical data acquired from Mt. John University Observatory (MJUO), used in this dissertation. Specific information regarding MJUO, and telescope settings and rig design, are also given. Lastly, a list of symbols used in Tables C.2 C.3, and C.4 is also provided.

Mt. John University Observatory:

Latitude:	-43 Degrees, 59.2 Minutes
Longitude:	170 Degrees, 27.9 Minutes East
Time Zone:	+12 hours GMT
Altitude:	1027 m

Telescope and wavefront sensor:

Telescope:	McLellan, 1-m, Dall-Kirkman Cassegrain reflector
Objective	1 m
Focal ratio:	f/7.7 and f/13.5 (f/7.7 used for data acquisition)
Rig:	Modified SCIDAR optical breadboard [63]
Wavefront sensor:	Curvature and Geometric [20]
CCD Cameras:	DragonFly Express (IEEE-1394b), Point Grey Research Inc.

List of symbols:

RA, Dec.	Right Ascension, Declination
UT	Universal time
HA	Hour angle
θ_{Est}	Estimated angular separation between objects (μrad)
T	Temperature (°C)
E	Exposure time (ms)
G	CCD Gain (dB)
ζ	Zenith angle (Deg.)
$V_{\mathrm{Wind}}(0)$	Wind speed at ground layer ($\mathrm{m\,s^{-1}}$)

Table C.1 Summary of observation runs at MJUO.

Run Date	Objects of interest	Weather and seeing conditions	Focal ratio	Camera sensor & Rig
26/10/05	μ Sco	Variable conditions	f/7.7	Micopix, SCIDAR
28/04/06	α Cen, μ Sco	Stable, good seeing	f/13.5	Micopix, SCIDAR
02/07/06	μ Cru, μ Sco, α Cen	Variable conditions	f/13.5	Micopix, SCIDAR
26/08/06	α Can, θ^1 Ori, α Cen	Good seeing	f/13.5	Micopix, SCIDAR
07/12/06	α Can, β Tuc	Poor seeing	f/7.7	Micopix, SCIDAR
26/01/07	α Can, α Cen	Good seeing	Various	Express, SCIDAR
15/06/07	α Cru, αCen, θSco	very good seeing	f/7.7	Express, Mod. SCIDAR
30/08/07	μ Cru, α Eri, β Cru	Variable conditions	f/7.7	Express, Mod. SCIDAR
29/10/08	Jovian moons	Good seeing	f/13.5	Express, Mod. SCIDAR
30/10/08	α Cen, θ^1 Ori	Good seeing	f/13.5	Express, Mod. SCIDAR
27/07/09	Jovian moons	Variable seeing	f/7.7	Express, Mod. SCIDAR
28/07/09	Jovian moons	Excellent seeing	f/7.7	Express, Mod. SCIDAR

Table C.2 Taylor's hypothesis experiment at coordinates RA: 21:48:49 and Dec: -14:24:30, on 27 July 2009.

Run ID	Object	Frames	UT	HA	Air Mass	θ_{Est} μrad	T (°C)	E (ms)	G (dB)	ζ (Deg)	$V_{\mathrm{Wind}}(0)$ ($\mathrm{m\,s^{-1}}$)
JUP_2a27a	Io, Europa	500	12:05	-01:58:44	1.2822	240	4.6	16.7	30	51.41	3.7
JUP_2a27b	Io, Europa	300	12:09	-01:55:05	1.2746	238	4.6	16.7	25	52.03	2.2
JUP_2a27c	Io, Europa	1000	12:33	-01:32:13	1.2275	180	4.8	16.7	20	54.57	5.0
JUP_2a27d	Io, Europa	1000	12:40	-01:20:39	1.2201	173	4.6	16.7	22	55.05	5.2
JUP_2a27e	Io, Europa	2000	12:48	-01:16:12	1.2113	166	4.6	16.7	22	56.81	5.2
JUP_2a27f	Io, Europa	1000	12:58	-01:11:31	1.1925	160	4.6	16.7	22	57.01	3.0
JUP_2a27g	Io, Europa	1000	12:58	-01:06:25	1.1878	152	4.8	16.7	22	57.35	2.0
JUP_2a27h	Io. Europa	10000	13:45	-00:24:22	1.1545	90	5.0	16.7	22	59.95	2.0
JUP_2a27i	Io, Europa	5000	13:59	-00:07:01	1.3893	49	5.0	16.7	22	45.95	2.0

Table C.3 Taylor's hypothesis experiment at coordinates RA: 21:48:46 and Dec: -14:25:10 on 28 July 2009.

Run ID	Object	Frames	UT	HA	Air Mass	θ_{Est} μrad	T (°C)	E (ms)	G (dB)	ζ (Deg)	$V_{\text{Wind}}(0)$ ($\text{m}\,\text{s}^{-1}$)
JUP_2a28	Europa, Ganymede	500	11:33	-02:27:29	1.7632	790	6.0	16.7	22	47.21	3.8
JUP_2a28a	Europa, Ganymede	100	11:40	-02:10:12	1.5434	730	6.0	16.7	22	49.24	2.0
JUP_2a28b	Europa, Ganymede	5000	12:04	-01:57:15	1.2801	700	6.0	16.7	22	51.42	3.6
JUP_2a28c	Europa, Ganymede	5000	12:30	-01:30:39	1.2241	640	6.3	16.7	22	54.86	3.6
JUP_2a28d	Europa, Ganymede	2000	12:49	-01:12:13	1.1957	260	6.3	16.7	22	56.79	3.4
JUP_2a28e	Europa, Ganymede	5000	13:42	-00:18:47	1.1671	320	6.2	16.7	22	60.26	3.6
JUP_2a28f	Io, Europa	5000	14:46	+00:45:46	1.1043	410	6.2	16.7	22	58.91	2.0
JUP_2a28g	Io. Europa	5000	15:01	+01:22:41	1.1854	420	6.3	16.7	22	55.65	2.0
JUP_2a28h	Europa, Ganymede	5000	15:55	+02:02:06	1.2895	320	6.4	16.7	22	51.32	1.5
JUP_2a28i	Io, Europa	5000	16:02	+02:02:07	1.2892	560	6.4	16.7	22	51.28	1.5
JUP_2a28j	Callisto	5000	16:12	+02:12:38	1.3890	-	6.3	16.7	22	49.14	2.0

Table C.4 Single and multiple star wavefront data used for time-series analysis and ESN training, 15 June, 2007.

Run ID	Object	RA	Dec.	Frames	UT	HA	Air Mass	T (°C)	E (ms)	G (dB)	ζ (Deg)	$V_{\text{Wind}}(0)$ ($\text{m}\,\text{s}^{-1}$)
NGS6a15_1	α cru	12:26:51	-63:08:30	5000	11:29	+04:02:40	1.284	3.2	16.7	20	50.93	3.5
NGS6a15_2	α cru	12:26:51	-63:12:20	5000	11:43	+04:11:50	1.312	3.2	16.7	15	49.6	5.0
NGS6a15_3	α cru	12:26:51	-63:12:20	1000	11:45	+04:20:20	1.321	3.2	16.7	10	49.11	2.5
NGS6a15_4	α cru	12:26:51	-63:12:20	5000	11:54	+04:27:30	1.337	3.0	16.7	08	48.2	2.0
NGS6a15_5	α cru	12:26:51	-63:12:20	5000	12:45	+05:18:47	1.474	3.0	16.7	10	42.63	2.0
NGS6a15_6	α cen	13:16:01	-60:54:49	5000	13:16	+03:36:43	1.232	3.2	16.7	4	54.18	3.0
NGS6a15_7	α cen	13:25:00	-60:54:49	5000	13:30	-	-	-	16.7	4	-	3.5
NGS6a15_8	θ sco	17:37:55	-43:00:09	5000	14:12	+01:34:44	1.047	2.4	16.7	20	72.78	3.8
NGS6a15_9	θ sco	17:37:55	-43:00:09	5000	14:23	+01:45:42	1.058	2.4	16.7	20	70.82	3.8
NGS6a15_10	θ sco	17:37:55	-43:00:09	5000	14:44	+02:10:34	1.095	2.4	16.7	20	66.4	3.8
NGS6a15_11	θ sco	17:37:55	-43:00:09	5000	15:00	+02:22:21	1.110	2.5	16.7	20	66.2	3.8
NGS6a15_12	θ sco	17:37:55	-43:00:09	1000	15:26	+02:49:24	1.159	1.0	16.7	15	59.54	3.8
NGS6a15_13	θ sco	17:37:55	-43:00:09	5000	15:40	+03:02:44	1.188	1.0	16.7	20	57.22	3.8

Appendix D

Derivation of the Continuous and Discrete Point Spread Function

The point spread function (PSF) is an an important quantity used throughout this dissertation. As shown in Subsection 3.7.2, the spatially variant characterisation of the PSF is used for anisoplanatic imaging. The derivation of the PSF provides insight into the foundations of the imaging model. The discrete version is also derived in this section.

Firstly, a circular pupil function $P(\boldsymbol{x}_1)$ is defined in the pupil plane, with diameter d,

$$P(\boldsymbol{x}_1) = P(x_1, y_1) = \text{rect}\left[\frac{d}{2}\left(x_1^2 + y_1^2\right)^{1/2}\right]. \tag{D.1}$$

A planar wavefront, incident on a lens or parabolic mirror, results in a spherical wavefront

$$U(R) = R^{-1}\exp(-jkR), \tag{D.2}$$

where $k = 2\pi/\lambda$, and R is the distance from a lens or mirror to the focus point on the image plane, $\mathbf{x}_2$. At the image plane, Equation D.2 is written in terms of Huygens-Fresnel diffraction integral

$$U_2(\mathbf{x}_2) = \frac{1}{j\lambda}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(\mathbf{x}_1)\frac{1}{R}\exp(-jkR)\frac{1}{r}\exp(jkr)\cos(\theta)\,\mathrm{d}\mathbf{x}_1, \tag{D.3}$$

where $P(\mathbf{x}_1)$ is the aperture pupil function, s_i is the shortest propagation distance from the lens to a focal point, and where $R = [x_1^2 + y_1^2 + z_1^2]^{1/2}$, and $r = [(x_2 - x_1)^2 + (y_2 - y_1)^2 + s_i^2]^{1/2}$.

In Equation D.3, θ can be assumed to be small, such that $\cos(\theta) \approx 1$. Also, R and r can be written in terms of s_i. Therefore, using the Fresnel approximation

$$R = s_i \left[1 + \frac{x_1^2}{s_i^2} + \frac{y_1^2}{s_i^2} \right]^{1/2}, \tag{D.4}$$

and

$$r = s_i \left[1 + \frac{(x_2 - x_1)^2}{s_i^2} + \frac{(y_2 - y_1)^2}{s_i^2} \right]^{1/2}. \tag{D.5}$$

The binomial expansion can be applied to Equations D.4 and D.5, such that

$$(1+x)^{1/2} = 1 + \frac{1}{2}x + \frac{1}{8}x^2 + \cdots, \tag{D.6}$$

when $|x| \ll 1$.

The binomial expansion of Equations D.4 and D.5 results in

$$R \approx s_i \left[1 + \frac{x_1^2}{2s_1^2} + \frac{y_1^2}{2s_1^2} \right], \tag{D.7}$$

and

$$r \approx s_i \left[1 + \frac{(x_2 - x_1)^2}{2s_i^2} + \frac{(y_2 - y_1)^2}{2s_i^2} \right], \tag{D.8}$$

respectively.

The field intensity at the image plane can now be approximated by applying Equations D.7 and D.8 to Equation D.3,

$$\begin{aligned} U_2(\mathbf{x}_2) = & \frac{1}{j\lambda s_i^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\mathbf{x}_1) \frac{1}{R} \exp \left[-jks_i \left(1 + \frac{x_1^2}{2s_1^2} + \frac{y_1^2}{2s_1^2} \right) \right] \\ & \times \exp \left\{ jks_i \left[1 + \frac{(x_2 - x_1)^2}{2s_i^2} + \frac{(y_2 - y_1)^2}{2s_i^2} \right] \right\} \mathrm{d}\mathbf{x}_1. \end{aligned} \tag{D.9}$$

Expanding the exponents, and collecting terms leads to

$$U_2(\mathbf{x}_2) = \frac{\exp \left[\frac{jk}{s_i^2}(x_1^2 + y_1^2) \right]}{j\lambda s_i^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\mathbf{x}_1) \exp \left[-j \frac{2\pi}{\lambda s_i} (x_2 x_1 + y_2 y_1) \right] \mathrm{d}\mathbf{x}_1. \tag{D.10}$$

By letting $u = x_1/\lambda s_i$ and $v = y_1/\lambda s_i$, Equation D.10 can be written as

$$U_2(\mathbf{x}_2) = -j\lambda \exp\left[\frac{jk}{s_i^2}(x_1^2+y_1^2)\right]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(\lambda s_i u, \lambda s_i v) \exp\big[-j2\pi(x_2 u + y_2 v)\big]\mathrm{d}u\mathrm{d}v. \tag{D.11}$$

Thus, Equation D.11 is the complex amplitude distribution of the field in the image plane, resulting from a point source at infinity in object space, and forms a focus known as the Fraunhofer diffraction pattern. The point spread function (PSF), $h(\mathbf{x}_2)$ comprises the integral portion of Equation D.11, and is the Fourier transform of the pupil function

$$h(\mathbf{x}_2) = \mathrm{FT}\big\{P(\lambda s_i u, \lambda s_i v)\big\}. \tag{D.12}$$

Considering the modified scaling factor, s_i, Equation D.12 is in agreement with Equation 3.9.

For isoplanatic regions, the spatially invariant PSF (SIPSF) is used to define the intensity distribution in the image plane. Thus, for incoherent light, the spatially invariant system response function is represented as a convolution

$$I_2(\mathbf{x}_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \|h(\mathbf{x}_2 - \mathbf{x}_1)\|^2 I_2(m\mathbf{x}_1)\mathrm{d}\mathbf{x}_1, \tag{D.13}$$

where m is the system magnification term.

In Equation D.13, the incoherent PSF represents the power spectrum of the pupil function. Since a generalised pupil function, $\mathscr{P}(\cdot)$, includes phase aberrations, Equation D.12 can be rewritten as

$$h_i(\mathbf{x}_2) = \big\|\mathrm{FT}\big\{\mathscr{P}(\lambda s_i u, \lambda s_i v)\big\}\big\|^2. \tag{D.14}$$

From Equations D.14 and 3.7, the incoherent PSF can be represented in rectangular coordinates as

$$h_i(x_2, y_2) = \frac{A^2}{\lambda^2 d^2}\left\|\mathrm{FT}\left\{P(x_1, y_1)\exp\left[-j\frac{2\pi}{\lambda}W(x_1, y_1)\right]\right\}\right\|^2, \tag{D.15}$$

where A is the aperture function, d is the distance from the exit pupil to the image plane, and $W(\cdot)$ is the wavefront phase function.

From Equation D.15, the discrete PSF is defined as

$$h_{id}(k,l) = \frac{A^2}{\lambda^2 d^2}\left\| \frac{1}{N^2}\sum_{n=0}^{\eta N-1}\sum_{m=0}^{\eta N-1} P_d(m.n) \times \exp\left[-j\frac{2\pi}{\eta N}(nk+ml)\right]\right\|^2, \tag{D.16}$$

where $k = x_2\frac{\eta A}{\lambda d}$, $l = y_2\frac{\eta A}{\lambda d}$, and where $x_1 = n\frac{A}{N}$, $y_1 = m\frac{A}{N}$, for integers, k, l.

Lucke provides a comprehensive discussion on the relationship between the continuous and discrete point spread function [79].

Appendix E

Summary of Mt. John University Observatory Wind Velocity Models

The following parameter values have been derived by Mohr [91], and form part of a larger study on atmospheric trending for Mt. John University Observatory. The values listed in Table E.1 are used in Chapter 4.

The following list defines symbols used in Table E.1.

List of symbols:

$V(H_T)$	Wind velocity at the tropopause
L_T	Thickness of the tropopause layer
H_T	Height of the tropopause layer
H_1	Height of the low altitude layer
L_1	Thickness of the low altitude layer
f_G	Greenwood frequency
$V(H_T)$	Wind speed of the tropopause layer ($\mathrm{m\,s^{-1}}$)
$V(0)$	Wind speed at ground layer ($\mathrm{m\,s^{-1}}$)

Table E.1 Parameters for wind velocity $V(h)$ models. f_G values are specified for $\lambda = 589$ nm [89].

Model	Usage	$V(0)$ ($\mathrm{m\,s^{-1}}$)	$V(H_T)$ ($\mathrm{m\,s^{-1}}$)	H_T (km)	L_1 (km)	$V(H_1)$ ($\mathrm{m\,s^{-1}}$)	H_1 (km)	L_1 (km)	$f_G(C_n^2(h)$:HV 5-7) (Hz)	$f_G(C_n^2(h)$:MJUO3) (Hz)
Bufton	—	5	30	9.8	4.8	0			29.5	40.6
MJUO1V	Very calm & clear	2	12	11	4.8	0			19.9	32.0
MJUO2V	Low ground winds	2	30	11	4.8	0			36.5	63.6
MJUO3V	Moderate ground winds	2	30	11	4.8	8	2.5	2	46.4	76.9
MJUO4V	High ground winds	2	30	11	4.8	20	2.5	2	65.9	105.0

References

[1] T. Albert. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2004.

[2] H. C. Andrews and B. R. Hunt. *Digital Image Restoration*. Prentice-Hall, 1977.

[3] M. Aubailly. *Reconstruction of anisoplanatic adaptive optics images*. PhD thesis, Dept. of Electrical Engineering, Michigan Technological University, 2005.

[4] M. Aubailly, M. C. Roggemann, and T. J. Schulz. Approach for reconstructing anisoplanatic adaptive optics images. *Applied Optics*, 46(24):6055–6063, 2005.

[5] M. Aubailly, M. C. Roggemann, and T. J. Schulz. Prediction of the space-varying point spread function for reconstruction of anisoplanatic adaptive optics images. In R.K. Tyson and M. Lloyd-Hart, editors, *Astronomical Adaptive Optics Systems and Applications II*, volume 5903, 590508. SPIE - The International society for Optical Engineering, 2005.

[6] Mathieu Aubailly and Michael C. Roggemann. Block-processing method for post-detection correction of anisoplanatic adaptive optics images. In *Adaptive Optics: Analysis and Methods/Computational Optical Sensing and Imaging/Information Photonics/Signal Recovery and Synthesis Topical Meetings on CD-ROM*, page JTuC4. Optical Society of America, 2005.

[7] G. R. Ayers and J. C. Dainty. Interative blind deconvolution method and its applications. *Opt. Lett.*, 13(7):547–549, 1988.

[8] M. R. Banham and A. K. Katsaggelos. Digital image restoration. *Signal Processing Magazine, IEEE*, 14(2):24–41, Mar 1997.

[9] V. Barakat, B. Guilpart, R. Goutte, and R. Prost. Model-based Tikhonov regularization and performance for a shift-varying degradation. *Applications of Digital Image Processing XX*, 3164(1):210–220, 1997.

[10] J. W. Beletic, R. J. Dorn, T. Craven-Bartle, and B. Burke. A new CCD designed for curvature wavefront sensing. In P. Amico and J. W. Beletic, editors, *Astrophysics and Space Science Library*, volume 252 of *Astrophysics and Space Science Library*, pages 283–308, 2000.

[11] R. Berry and J. Burnell. *The handbook of astronomical image processing*. Willmann-Bell Inc., 2001.

[12] B. B. Boone. *Signal Processing using Optics: Fundamentals, Devices, Architectures, and Applications*. Oxford University Press, 1996.

[13] M. C. Britton. The anisoplanatic point-spread function in adaptive optics. *Pub. Astron. Soc. Pac.*, 118:885–900, 2006.

[14] R. G. Brown and P. Y. C. Hwang. *Introduction to Random Signals and Applied Kalman Filtering*. John Wiley and Sons, 1996.

[15] L. Cao. Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D*, 110:43–50, 1997.

[16] M. Carbillet, C. Vérinaud, B. Femenía, A. Riccardi, and L. Fini. Modelling astronomical adaptive optics - I. The software package CAOS. *Royal Astronomical Society, Monthly Notices*, 356:1263–1275, FEB 2005.

[17] C. Chatfield. *The analysis of time series - An introduction*. Chapman and Hall/CRC, 6 edition, 2004.

[18] T. A. Cheema, I. M. Qureshi, and A. Hussain. Blind image deconvolution using space-variant neural network approach. *Electronics Letters*, 41(6):308–309, March 2005.

[19] T. Y. Chew. *Wavefront sensors in Adaptive Optics*. PhD thesis, Dept. of Electrical & Computer Engineering, University of Canterbury, 2008.

[20] T. Y. Chew and R. G. Lane. Estimating phase aberrations from intensity data. *Image and Vision Computing NZ 2003 (IVCNZ'03)*, 2003.

[21] M. Chun. The useful field of view of an adaptive optics system. *Astron. Soc. Pac.*, 110:317329, 1998.

[22] G. S. Chundi, M. Lloyd-Hart, and M. K. Sundareshan. Training multilayer perceptron and radial basis function neural networks for wavefront sensing and restoration of turbulence-degraded imagery. In *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*, volume 3, pages 2117 – 22, Budapest, Hungary, 2004.

[23] D. Dayton and S/ Sandven. Expectation maximiization approach to deconvolution from wavefront sensing. In T.J Schulz, editor, *Image Reconstruction and Restoration II*, volume 3170, pages 16–24. SPIE - The International society for Optical Engineering, 1997.

[24] N. J. Doelman, K J. G. Hinnen, F. J. G. Stoffelen, and M. H. Verhaegen. Optimal control strategy to reduce the temporal wavefront error in ao systems. In D. B. Calia, B. L. Ellerbroek, and R. Ragozzoni, editors, *Proceedings of SPIE Annual Conference on Adaptive Optical System Technologies*, volume 5490, pages 1426–1437. SPIE - The International society for Optical Engineering, 2004.

[25] R. W. Duffner. *The Adaptive Optics evolution - a history*. University of New Mexico Press, 2009.

[26] S. Engelberg. *Random Signals and Noise - a Mathematical Introduction*. CRC Press, 2007.

[27] H. Ferguson. Space-based vs. ground-based telescopes with adaptive optics (AO), 2002. Space Telescope Science Institute, Association of Universities for Research in Astronomy (AURA).

[28] D. Fraser, A. Lambert, M. R. S. Jahromi, M. Tahtali, and D. Clyde. Anisoplanatic image restoration at ADFA. In *Proceedings of the seventh international conference on digital image computing, techniques and applications, DICTA, 2003*, pages 19 – 27, Sydney, Australia, 2003.

[29] D. L. FRIED. Optical resolution through a randomly inhomogeneous medium for very long and very short exposures. *J. Opt. Soc. Am.*, 56(10):1372–1379, 1966.

[30] D. L. Fried. Anisoplanatism in adaptive optics. *J. Opt. Soc. Am.*, 72:52 – 61, 1982.

[31] D. L. Fried. Post detection wavefront compensation. In *Proceedings of SPIE conference on digital image recovery and synthesis*, pages 128 – 133. SPIE, 1987.

[32] T. Fusco, J.-M. Conan, V. Michau, L. M. Mugnier, and G. Rousset. Efficient phase estimation for large-field-of-view adaptive optics. *Optical Letters*, 24(21):1472–1474, 1999.

[33] T. Fusco, J.-M. Conan, L. M. Mugnier, V. Michau, and G. Rousset. Characterization of adaptive optics point spread function for anisoplanatic imaging. Application to stellar field deconvolution. *Astronomy and Astrophysics*, 142:149–156, February 2000.

[34] P. J. Gallant and J. K. Aitken. Genetic algorithm design of complexity-controlled time-series predictors. *IEEE XIII Workshop on Neural Networks for Signal Processing*, pages 769–778, 2003.

[35] J. D. Gaskill. *Linear Systems, Fourier Transforms, and Optics*. Wiley series in pure and applied optics, 1978.

[36] J. M. Geary. Wavefront sensors. In R. K. Tyson, editor, *Adaptive Optics Engineering Handbook*, pages 123–150. M. Dekker, 2000.

[37] E. Gendron and P. Léna. Single Layer Atmospheric Turbulence Demonstrated by Adaptive Optics Observations. *Astrophysics and Space Science*, 239:221–228, sep 1996.

[38] R. W. Gerchberg and O.W. Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237–247, 1976.

[39] R. A. Gonsalves. Phase retrieval from modulus data. *J. Opt. Soc. Am.*, 66(9):961–964, 1976.

[40] R. C. Gonzalez and R E. Woods. *Digital Image Processing*. Addison-Wesley Publishing Company Inc., 1992.

[41] I. J. Good. What are degrees of freedom? *The American Statistician*, 27(5):227–228, 1973.

[42] J. W. Goodman. *Introduction to Fourier Optics*. McGraw-Hill, 1968.

[43] J. W. Goodman. *Statistical Optics*. McGraw-Hill, 1985.

[44] J. W. Goodman. *Introduction to Fourier Optics*. McGraw-Hill, third edition, 2005.

[45] D. P. Greenwood. Bandwidth specification for adaptive optics systems. *J. Opt. Soc. Am.*, 67(3):390–393, 1977.

[46] P. C. Hansen. Deconvolution and regularization with Toeplitz matrices. *Numerical Algorithms*, 29:323–378, 2002.

[47] C. M. Harding, R. A. Johnston, and R. G. Lane. Fast simulation of a Kolmogorov phase screen. *App. Opt.*, 38(11):2161–2170, 1999.

[48] J. W. Hardy. Active optics: A new technology for the control of light. *Proc. IEEE*, 66(6):651–697, 1978.

[49] J. W. Hardy. *Adaptive optics for astronomical telescopes*. Oxford University Press., New York, NY, USA, 1998.

[50] S. Haykin and J. Principe. Using neural networks to dynamically model chaotic events such as sea clutter. In IEEE Press, editor, *IEEE Signal Processing Magazine*, pages 66–81, May 1998.

[51] J. Hertz, A. Krogh, and R. G. Palmer. *Introduction to the theory of neural computation: Lecture notes Volume I*. Westview Press, 1991.

[52] P. Hickson. Wave-front curvature sensing from a single defocused image. *J. Opt. Soc. Am. A*, 11(5):1667–1673, 1994.

[53] S. B. Howell. *Handbook of CCD astronomy*. Cambridge University Press., Cambridge, United Kingdom, 2000.

[54] B.R. Hunt, D.G. Sheppard, and M.S. Nadar. Multiframe poisson map deconvolution of astronomical images. In *Image Processing, 1996. Proceedings., International Conference on*, volume 3, pages 109 –111 vol.3, 16-19 1996.

[55] The Mathworks Inc. Technical solutions, July 2006.

[56] T. Jabara. *Machine learning: discriminative and generative*. Kluwer Academic Publishers, 2004.

[57] H. Jaeger. The echo state approach to analysing and training recurrent neural networks. Technical Report GMD Report 148, German National Research Center for Information Technology, 2001.

[58] H. Jaeger. Short term memory in echo state networks. Technical Report GMD Report 152, Fraunhofer Institute for Autonomous Intelligent Systems, March 2002.

[59] H. Jaeger and H. Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304:78–80, 2001.

[60] H. Jaeger, M. Lukoševičius, D. Popovici, and U. Siewert. Optimization and applications of echo state networks with leaky- integrator neurons. *Neural Networks*, 20(3):335–352, 2007.

[61] H. Jakobsson. Simulations of time series of atmospherically distorted wave fronts. *Appl. Opt.*, 35(9):1561–1565, 1996.

[62] R. Johnston. *Inverse Problems in Astronomical Imaging*. PhD thesis, Dept. of Electrical & Computer Engineering, University of Canterbury, 2001.

[63] R . A. Johnston, J. L. Mohr, P. L. Cottrell, and R. G. Lane. A bread-board SCIDAR system at Mount John. In P. J. Bones, editor, *in proceedings of Image & Vision Computing NZ 2007 (IVCNZ'07)*, Christchurch, Dec 2004.

[64] R. A. Johnston and R. G. Lane. Modeling scintillation from an aperiodic Kolmogorov phase screen. *Applied Optics*, 39(26):4761–4769, 2000.

[65] M. B. Jorgenson and G. J. M. Aitken. Evidence of a chaotic attractor in star-wander data. *Optics Letters*, 16(2), 1991.

[66] M. B. Jorgenson and G. J. M. Aitken. Wavefront Prediction for Adaptive Optics. In F. Merkle, editor, *European Southern Observatory Astrophysics Symposia*, volume 48 of *European Southern Observatory Astrophysics Symposia*, pages 143–+, January 1994.

[67] B. L. Kalman and s. C. Kwasny. Why tanh: choosing a sigmoidal function. *Neural Networks, 1992. IJCNN., International Joint Conference on*, 4:578–581, Jun 1992.

[68] A. K. Katsaggelos. Iterative image restoration algorithms. *Optical Engineering*, 28:735–748, jul 1989.

[69] A. K. Katsaggelos. Introduction. In Aggelos Konstantinos Katsaggelos, editor, *Digital Image Restoration*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1991.

[70] O. Keskin, P. Hampton, R. Conan, C. Bradley, A. Hilton, and C. Blain. Woofer-Tweeter adaptive optics test bench. In *Adaptive Hardware and Systems, 2006. AHS 2006. First NASA/ESA Conference on*, pages 74–80, June 2006.

[71] O. Keskin, J. Laurent, and B. Colin. Hot-air optical turbulence generator for the testing of adaptive optics systems: principles and characterization. *Appl. Opt.*, 45(20):4888–4897, 2006.

[72] A. N. Kolmogorov. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Proc. R. Soc. Lond.*, 434(1890):9–13, 1991.

[73] M. Korenberg, D. McGaughey, and G. J. M. Aitken. Parallel cascade prediction of turbulence induced wavefront tilt. *Electronics Letters*, 32(14):1315–1316, Jul 1996.

[74] M. Le Louarn. Multi-Conjugate Adaptive Optics: a PSF study. In E. Vernet, R. Ragazzoni, S. Esposito, and N. Hubin, editors, *European Southern Observatory Astrophysics Symposia*, volume 58 of *European Southern Observatory Astrophysics Symposia*, pages 217–222, 2002.

[75] W. Y. V. Leung. *Inverse problems in astronomical and general imaging*. PhD thesis, Dept. of Electrical & Computer Engineering, University of Canterbury, 2002.

[76] M. Lloyd-Hart and P. McGuire. Spatio-temporal prediction for adaptive optics wavefront reconstructors. *Proceedings ESO*, pages 95–101, 1995.

[77] Michael Lloyd-Hart, Christoph Baranec, N. Mark Milton, Miguel Snyder, Thomas Stalcup, and J. Roger P. Angel. Experimental results of ground-layer and tomographic wavefront reconstruction from multiple laser guide stars. *Opt. Express*, 14(17):7541–7551, 2006.

[78] A. W. Lohmann, G. Weigelt, and B. Wirnitzer. Speckle masking in astronomy: triple correlation theory and applications. *Appl. Opt.*, 22(24):4028–4037, 1983.

[79] R. L. Lucke. The relation between physical and computer-generated point spread functions and optical transfer functions. *American Journal of Physics*, 69:1237–1244, December 2001.

[80] M. Lukoševičius and H. Jaeger. Overview of reservoir recipes. Technical Report 11, School of Engineering and Science, Jacobs University Bremen, 2007.

[81] M. Lukoševičius and H. Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, August 2009.

[82] W. Maass, T. Natschläger, and H. Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 14(11):2531–2560, 2002.

[83] P-Y. Madec, J-M. Conan, and G. Rousset. Temporal characterization of atmospheric wavefront for adaptive optics. In M. H. Ulrich, editor, *Proceedings of European*

Southern Observatory, Conference 42 on progress in telescope and instrumentation technologies, pages 471–474, 1992.

[84] V. N. Mahajan. *Optical imaging and aberrations, Part 1: Ray Geometrical Optics*. The International Society for Optical Engineering - SPIE., Washington, USA, 1998.

[85] D. P. Mandic and J. Chambers. *Recurrent Neural Networks for Prediction: Learning Algorithms,Architectures and Stability*. John Wiley & Sons, Inc., New York, NY, USA, 2001.

[86] J. M. Martin and S. M. Flatté. Intensity images and statistics from numerical simulation of wave propagation in 3-D random media. *Appl. Opt.*, 27(11):2111–2126, 1988.

[87] T. Masters. *Practical Neural Network Recipes in C++*. Morgan Kaufmann Publishers, 1993.

[88] D. R. McGaughey and G. J. M. Aitken. Temporal analysis of stellar wave-front-tilt data. *J. Opt. Soc. Am. A*, 14(8):1967–1974, 1997.

[89] C. D. McGillem and G. R. Cooper. *Continuous and Discrete Signal and System Analysis*. Saunders College Publishing, 3 edition, 1991.

[90] P. C. McGuire, D. G. Sandler, M. Lloyd-Hart, and T. A. Rhoadarmer. Adaptive optics: Neural network wavefront sensing, reconstruction, and prediction. *Scientific Applications of Neural Nets, Proceedings of the 194th W. E. Heracus Seminar, 1998. Edited by J.W. Clark, T Lindenau, and M.L. Ristig, Springer-Verlag Publishers*, page 97, 1999.

[91] J. L. Mohr. *Atmospheric turbulence characterisation using scintillation detection and ranging*. PhD thesis, Dept. of Physics & Astronomy, University of Canterbury, 2009.

[92] J. L. Mohr, R. A. Johnston, C. C. Worley, and P. L. Cottrell. Optical turbulence profling at Mount John University Observatory. In A. Kohnle, K. Stein, and J. D. Gonglewski, editors, *Optics in Atmospheric Propagation and Adaptive Systems XI*, volume 7108, page 70809, 2008.

[93] R. Molina, J. Nunez, F. J. Cortijo, and J. Mateos. Image restoration in astronomy: a Bayesian perspective. *Signal Processing Magazine, IEEE*, 18(2):11–29, Mar 2001.

[94] D. A. Montera, B. M. Welsh, M. C. Roggemann, and D. W. Ruck. Prediction of wave-front sensor slope measurements with artificial neural networks. *Appl. Opt.*, 36(3):675–681, 1997.

[95] L Mortara and A Fowler. Solid state images for astronomy. In *Proceedings of SPIE*, volume 290 of *28*. SPIE - The International society for Optical Engineering, 1981.

[96] L. Mugnier, G. Le Besnerais, and S. Meimon. Inversion in optical imaging through atmospheric turbulence. In J. Idier, editor, *Bayesian approach to inverse problems*, pages 257–258. John Wiley & Sons Ltd., 2008.

[97] HM Nautical Almanac Office. *The Astronomical Almanac, 2009*. Her Majesty's Nautical Almanac Office, 2009.

[98] R. J. Noll. Zernike polynomials and atmospheric turbulence. *Opt. Soc. Am.*, 66:207–211, 1976.

[99] Dept. of Electrical & Computer Engineering. Beowulf Cluster, 2009. University of Canterbury.

[100] University of Southern California Signal and Image Processing Institute. The usc-sipi image database. Website, 2008. `http://sipi.usc.edu/services/database/Database.html`.

[101] The Optical Sciences Company. AO Tools, 2008. `http://www.tosc.com`.

[102] M. Ozturk, D. Xu, and J. Principe. Analysis and design of echo state networks. *Neural Computation*, 19:111–138, 2007.

[103] R. R. Parenti and R. J. Sasiela. Laser-guide-star systems for astronomical applications. *J. Opt. Soc. Am. A*, 11(1):288–309, 1994.

[104] U. Parlitz, I. Wedekind, W. Lauterborn, and C. Merkwirth. TSTOOL and user manual, ver. 1.11. University of Göttingen, 2002.

[105] P. Piatrou and M. C. Roggemann. Performance study of Kalman filter controller for multiconjugate adaptive optics. *Appl. Opt.*, 46(9):1446–1455, 2007.

[106] J. Primot, G. Rousset, and J. C. Fontanella. Deconvolution from wave-front sensing: a new technique for compensating turbulence-degraded images. *J. Opt. Soc. Am. A*, 7(9):1598–1608, 1990.

[107] B. H. Pui, B. Hayes-Gill, M. Clark, M. Somekh, C. See, J. F. Pieri, S. Morgan, and A. Ng. The design and characterisation of an optical vlsi processor for real time centroid detection. In *Proceedings on the deisgn, test, integration, and packaging of MEMS/MOEMS*, volume v4408, pages 73–80. SPIE - The International society for Optical Engineering, 2001.

[108] R. Ragazzoni. No Laser Guide Stars for adaptive optics in giant telescopes? *Astron. Astrophys. Suppl. Ser.*, 136:205–209, April 1999.

[109] R. Ragazzoni, E. Marchetti, and F. Rigaut. Modal tomography for adaptive optics. *Astron. Astrophys.*, 342:L53–L56, 1999.

[110] R. Ragazzoni, E. Marchetti, and G. Valento. Adaptive-optics corrections available for the whole sky. *Nature (UK)*, 403(6765):54 – 6, 2000/01/06.

[111] J. Reményi, P. Várhegyi, L. Domján, P. Koppa, and E. L orincz. Amplitude, phase, and hybrid ternary modulation modes of a twisted-nematic liquid-crystal display at 400 nm. *Appl. Opt.*, 42(17):3428–3434, 2003.

[112] W. H. Richardson. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.*, 62(1):55–59, 1972.

[113] F. Roddier. Effects of atmospheric turbulence in astronomy. *Progress in optics*, 19:281–376, 1981.

[114] F. Roddier. Curvature sensing and compensation: a new concept in adaptive optics. *Applied Optics*, 27(7):1223–1225, 1988.

[115] F. Roddier. *Adaptive Optics in Astronomy*. Cambridge University Press, 1990.

[116] M. C. Roggemann, B. L. Ellerbroek, and T. A. Rhoadarmer. Widening the effective field of view of adaptive-optics telescopes by deconvolution from wave-front sensing: average and signal-to-noise ratio performance. *Appl. Opt.*, 34(8):1432–1444, 1995.

[117] M. C. Roggemann and B. Welsh. *Imaging Through Turbulence*. CRC Press, 1996.

[118] A. Rosenfeld and A. C. Kak. *Digital picture processing*. Academic Press Inc., 2 edition, 1982.

[119] W. S. Sarle. Stopped training and other remedies for overfitting. In P. V. Marsden, editor, *Proceedings of the 27th Symposium on Computing Science and Statistics*, pages 352–360, 1995.

[120] G. Schneider. Domains of observability in the near-infrared with HST/NICMOS and adaptive optics augmented large ground-based telescopes, 2002. A summary study solicited by STScI in preparation for HST Cycle 12, University of Arizona.

[121] B. Schrauwen, D. Verstraeten, and J. Van Campenhout. An overview of reservoir computing: theory, applications and implementations. In *Proceedings of the 15th European Symposium on Artificial Neural Networks*, pages 471–482, 2007.

[122] C. Schwartz, G. Baum, and E. N. Ribak. Turbulence-degraded wave fronts as fractal surfaces. *J. Opt. Soc. Am. A*, 11(1):444–451, 1994.

[123] J. C. Sprott. *Chaos and Time-Series Analysis*. Oxford University Press, 2003.

[124] G. Strang. *Linear algebra and its applications*. Harcourt College Publishers, 3rd edition, 1988.

[125] F. Takens. Detecting strange attractors in turbulence. In D. A. Rand and L. S. Young, editors, *Dynamical systems and turbulence, lecture notes in mathematics*, volume 898, pages 366–381. Springer-Verlag., 1981.

[126] M. Tallon and R. Foy. Adaptive telescope with laser probe - Isoplanatism and cone effect. *Astronomy and Astrophysics*, 235:549–557, Aug 1990.

[127] V. I. Tatarski. *Wave propagation in a turbulent medium*. McGraw-Hill Book Company, New York, NY, USA, 1961.

[128] G. I. Taylor. Statistical Theory of Turbulence. *Royal Society of London Proceedings Series A*, 151:421–444, September 1935.

[129] R. Y. T. Tiong and S. J. Weddell. A hot-air turbulence generator for optical test-benches. In *In the proceedings of the Electronics New Zealand Conference (ENZ-Con'07)*, pages 71–76, 2007.

[130] N. Tishby, F. C. Pereira, and W. Bialek. The information bottleneck method. In *In proceedings of the 37th annual Allerton conference on communication, control, and computing*, pages 368–377, 1999.

[131] A. Tokovinin, M. Le Louarn, E. Viard, N. Hubin, and R. Conan. Optimised modal tomograph in adaptive optiics. *Astronomy & Astrophysics*, 378:710–721, 2001.

[132] A. Tokovinin and E. Viard. Limited precision of tomographic phase. *Opt. Soc. Am.*, 18(4):873–881, April 2001.

[133] G. A. Tyler. Bandwidth considerations for tracking through turbulence. *J. Opt. Soc. Am. A*, 11(1):358–367, 1994.

[134] R. K. Tyson. *Introduction to Adaptive Optics*. Optical Engineering, IEEE Press, 2000.

[135] R. K. Tyson and B. W. Frazier. *Field guide to Adaptive Optics*. The International Society for Optical Engineering, SPIE Press, 2004.

[136] M. A. van Dam and R. G. Lane. Extended analysis of curvature sensing. *J. Opt. Soc. Am.*, 19(7):1390–1397, 2002.

[137] G. V. Vdovin. Model of an adaptive optical system controlled by a neural network. *Optical Engineering*, 34(11):3249–3253, 1995.

[138] J. Vernin and F. Roddier. Experimental determination of two-dimensional spatiotemporal power spectra of stellar light scintillation. evidence for a multilayer structure of the air turbulence in the upper troposphere. *J. Opt. Soc. Am.*, 63(3):270–273, 1973.

[139] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.

[140] S. J. Weddell. An embedded synchronisation module for enhanced imaging. Technical Report ECE Report SJW-3, University of Canterbury, Electrical and Computer Engineering, and Physics and Astronomy, 2005.

[141] S. J. Weddell and R. Y. Webb. Dynamic artificial neural networks for centroid prediction in astronomy. In IEEE Press, editor, *Sixth International Conference, Hybrid Intelligent Systems (HIS'06), Auckland, New Zealand*, pages 68–72. IEEE Press, Dec 2006.

[142] S. J. Weddell and R. Y. Webb. Reservoir computing for prediction of the spatially-variant point spread function. *Selected Topics in Signal Processing, IEEE Journal of*, 2(5):624–634, Oct. 2008.

[143] S. J. Weddell and R. Y. Webb. The restoration of extended astronomical images using the spatially-variant point spread function. In K. Irie, editor, *in proceedings of Image and Vision Computing New Zealand, 2008 (IVCNZ'08)*, pages 1–6. IEEE Press, Nov 2008.

[144] Zhiying Wen, Donald Fraser, and Andrew Lambert. Bicoherence: a new lucky region technique in anisoplanatic image restoration. *Appl. Opt.*, 48(32):6111–6119, 2009.

[145] R. W. Wilson. Slodar: Measuring optical turbulence altitude with a shack-hartmann wavefront sensor. *Royal Astronomical Society*, 337, 2002.

[146] A. Wolf, J. B. Swift, L. H. Swinney, and J. A. Vastano. Determining Lyapunov exponent from a time series. *Physica D*, 16:285–317, 1985.

[147] C. C. Worley. Atmospheric turbulence monitoring at Mt. John. Technical Report ASTR-480 Honors project, University of Canterbury, Physics and Astronomy, 2005.

[148] J. C. Wyant and K. Creath. *Basic wavefront aberration theory for optical metrology*, pages 9–15. Academic Press Inc., USA, 1992.

[149] D. Xu, J. Lan, and J. C. Principe. Direct adaptive control: an echo state network and genetic algorithm approach. In *Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on*, volume 3, pages 1483–1486, 2005.